

Article

Zero-Sum Game-Based Reinforcement Learning Tracking Control with Predefined-Time Prescribed Performance for Highly Flexible Aircraft

Hanwen Zhang ¹, Chi Peng ^{2,3,4}, Yuxin Zhang ¹, Jianjun Ma ¹  and Meiping Wu ^{1,*}

¹ College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410000, China; zhanghanwen19@alumni.nudt.edu.cn (H.Z.); zhangyuxin.24c@nudt.edu.cn (Y.Z.); mjj_nudt@nudt.edu.cn (J.M.)

² Xi'an Research Institute of Surveying and Mapping, Xi'an 710054, China; pengchi2729@nudt.edu.cn

³ State Key Laboratory of Spatial Datum, Xi'an 710054, China

⁴ National Key Laboratory of Intelligent Spatial Information, Beijing 100029, China

* Correspondence: wumeiping@nudt.edu.cn

Abstract

This paper develops a zero-sum game-based reinforcement learning tracking controller with predefined-time prescribed performance (ZG-RL-PP) for highly flexible aircraft. The disturbed tracking-error dynamics are first transformed into a min–max optimal control problem, where the control input and the disturbance are treated as two players with opposite objectives. To guarantee the prescribed transient and steady-state tracking performance, logarithmic barrier Lyapunov functions are incorporated into the value function and the Hamilton–Jacobi–Isaacs equation. For higher-relative-degree tracking-error channels, recursive auxiliary constraint variables are introduced to preserve the prescribed bounds on the original errors and enable constraint enforcement through the derivative channels in which the control inputs appear. A critic neural network is employed to approximate the value function online, and a predefined-time fractional-power learning law is adopted for critic weight updating. It is shown that the critic weight-estimation error is practically predefined-time convergent and that all closed-loop signals are uniformly ultimately bounded. Simulation results demonstrate the effectiveness and robustness of ZG-RL-PP in terms of tracking accuracy, disturbance attenuation, prescribed-performance satisfaction, and predefined-time learning.

Keywords: highly flexible aircraft; zero-sum game; reinforcement learning; tracking control; predefined-time prescribed performance; barrier function



Academic Editors: Davood Asadi and Hassan Haghighi

Received: 31 July 2026

Revised: 29 August 2026

Accepted: 2 September 2026

Published: 6 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Highly flexible aircrafts (HFAs) are representative benchmarks for advanced flight control due to their lightweight high-aspect-ratio wings and strong aeroelastic coupling. Nonlinear aeroelasticity and flight dynamics must be treated together for HFA aircraft since flexible deformation changes both the structural response and rigid-body motion [1]. Large elastic deformation in flight induces strong rigid–flexible coupling in HFA, resulting in pronounced nonlinearities and modeling uncertainties. These characteristics make controller design for HFA particularly challenging [2].

A variety of nonlinear control methods have been developed for flexible and very flexible aircraft. Gregory used modified nonlinear dynamic inversion to improve flexible-aircraft control, but this type of approach still relies on a carefully constructed dynamic

model [3]. Patil and Hodges studied output-feedback control for nonlinear aeroelastic response, showing that feedback design can suppress flexible response when the aeroelastic model is available [4]. Shearer and Cesnik proposed a trajectory-control architecture for very flexible aircraft, and Raghavan and Patil extended trajectory control ideas to high-aspect-ratio flying wings [5,6]. Gust and load-alleviation studies then introduced robust control, model predictive control, and LQG-based designs to improve performance under atmospheric excitation [7]. Cook et al. emphasized robust gust alleviation and stabilization for very flexible aircraft, while Haghighat et al. used model-predictive control for gust-load alleviation [8,9]. A state-of-the-art nonlinear robust control architecture is developed in [10] based on high-order sliding-mode estimation and disturbance reconstruction, which can effectively improve tracking performance under uncertain operating conditions. Liu et al. showed that LQG-based model-predictive control can be effective for gust-load alleviation in flexible aircraft [11]. However, Qu and Annaswamy pointed out that trim drift, unmeasured flexible modes, and actuator anomalies can exceed the robustness margin of conventional linear output-feedback designs, motivating adaptive output-feedback control with closed-loop reference models [12].

Reinforcement learning (RL) and approximate dynamic programming (ADP) provide another route because they learn value functions and policies from system data rather than requiring a completely known drift model. Lewis and Vrabie connected reinforcement learning and ADP to feedback control, and Lewis et al. later summarized how natural decision methods can be used to design optimal adaptive controllers [13,14]. Kiumarsi et al. reviewed optimal and autonomous control using reinforcement learning, showing the breadth of ADP/RL in modern feedback-control design [15]. For continuous-time systems, Vrabie et al. introduced policy-iteration-based adaptive optimal control, and Jiang and Jiang developed computational adaptive optimal control for systems with completely unknown dynamics [16,17]. Zhu et al. further showed that integral reinforcement learning can be used for suboptimal output-feedback control of partially unknown linear continuous-time systems [18]. Then, Peng and Ma used online integral RL to stabilize an uncertain HFA using full-state and output-feedback information, reducing the dependence on exact drift dynamics [2]. Nevertheless, RL controllers for highly flexible aircraft still face two important difficulties: the learning convergence time is not explicitly assigned by the designer, and external disturbances are not always treated as worst-case strategic inputs during the learning process.

Zero-sum differential games provide a natural framework for robust RL because the controller minimizes the performance index while the disturbance attempts to maximize it. Abu-Khalaf et al. used neurodynamic programming and zero-sum games to handle constrained control, which established an important link between Hamiltonian function equations and robust learning-based control [19]. Vamvoudakis and Lewis then developed an online solution for nonlinear two-player zero-sum games using synchronous policy iteration [20]. More recently, Wang et al. combined safe reinforcement learning, fixed-time stability, and zero-sum games to address nonlinear systems with external disturbances and obstacle-avoidance awareness [21]. However, fixed-time bounds can still be conservative, and the designer may not be able to set the actual convergence time directly as a mission parameter. To directly set the convergence time, predefined-time stability assigns an explicit convergence upper bound through a design parameter rather than through conservative constants that depend on initial conditions. Yuan et al. developed neural adaptive fixed-time control for nonlinear systems with full-state constraints, illustrating the value of neural approximation in time-constrained control [22]. However, zero-sum game based predefined-time reinforcement learning updated law has not been designed for highly flexible aircraft.

For HFA tracking, mission-level requirements also specify how large the transient tracking error is allowed to be before convergence and how accurate the steady-state response must be after convergence. Prescribed-performance control (PPC) was introduced to convert these engineering requirements into a time-varying inequality constraint on the tracking error. The classical PPC idea designs a positive performance function and an error transformation so that the constrained error evolves inside a designer-specified envelope, thereby assigning overshoot, convergence rate, and residual accuracy before the controller is synthesized [23,24]. Subsequent approximation-free and adaptive prescribed-performance schemes further showed that such envelope-based specifications can be incorporated into nonlinear feedback design without relying only on post hoc gain tuning [25].

Nevertheless, conventional PPC usually prescribes a decay profile rather than an exact physical time at which the final accuracy band must be reached. This is restrictive for flight missions because tracking transients are often scheduled by maneuver duration, actuator authority, and safety boundary. Prescribed-time and predefined-time control address this limitation by making the settling-time upper bound an explicit design parameter, so that the convergence requirement is fixed by the designer rather than inferred from conservative Lyapunov constants [26–28]. An effective flexible performance-based predefined-time tracking control is proposed for an unmanned helicopter with irregular constraints in [29], showing that predefined-time convergence and performance constraints can be handled simultaneously in aerial-vehicle control. Combining prescribed performance with prescribed time therefore gives a more direct mission-level specification: before the assigned time, the error is forced to remain inside a shrinking transient envelope, and after that time, it remains inside the prescribed steady-state accuracy band. Recent prescribed-time prescribed-performance designs also indicate that transient envelope shaping and finite-time accuracy assignment can be treated simultaneously for uncertain nonlinear systems [30–32]. However, a zero-sum game-based reinforcement learning tracking method predefined-time prescribed performance and predefined-time critic learning has not been fully developed for highly flexible aircraft.

In addition, the difficulty in the present HFA problem is that some constrained elastic-state channels are not directly actuated because the corresponding rows of the input matrix are zero. Standard PPC transformations or direct barrier penalties implicitly assume that the control input can react through the constrained output channel. For the HFA model, this assumption is not valid for high-relative-degree tracking errors. A direct logarithmic barrier on these channels cannot by itself transmit the boundary-keeping action to the control input. This motivates a barrier construction that combines direct boundary margins for actuated channels with recursive boundary margins for high-relative-degree channels. In this way, the prescribed-time performance requirement is still imposed on the original physical error, while the control action is generated through the derivative channel in which the input actually appears.

Motivated by the above discussions, this work develops a zero-sum game-based predefined-time prescribed-performance reinforcement learning method for robust tracking control of highly flexible aircraft. The main contributions are summarized as follows:

- Compared with existing reinforcement learning control for highly flexible aircraft [2], where disturbances are treated as exogenous signals, a zero-sum differential-game framework is developed to characterize the worst-case interaction between the controller and disturbance, converting robust tracking into a min–max optimal control problem.
- Compared with conventional zero-sum reinforcement learning methods [19,20] and fixed-time zero-sum reinforcement learning [21], predefined-time performance is introduced to explicitly constrain transient and steady-state tracking errors. The

resulting logarithmic barrier functions are incorporated into the value function and Hamilton–Jacobi–Isaacs equation, while recursive boundaries extend the performance constraints to higher-relative-degree tracking-error channels without direct actuation.

- A predefined-time critic learning law and stability analysis are established. The current-and-history fractional-power critic update law is retained to guarantee practical predefined-time convergence of the critic weight-estimation error under sufficient excitation. A composite Lyapunov analysis further proves the practical predefined-time convergence of the critic approximation error and the uniform ultimate boundedness of all closed-loop signals.

2. Problem Statement

2.1. Preliminaries

Consider the following autonomous nonlinear system:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)), \quad \mathbf{x}(0) = \mathbf{x}_0, \quad t \geq 0, \quad (1)$$

where $\mathbf{x}(t) \in \mathbb{R}^n$, $\mathbf{x}_0$ is the initial state, and $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is continuous with $\mathbf{f}(\mathbf{0}) = \mathbf{0}$.

Definition 1. If the equilibrium point $\mathbf{x} = \mathbf{0}$ is globally fixed-time stable and there exists a settling-time upper bound independent of the initial state, denoted by $T_{\max}$, and if this upper bound is constrained by a given constant $T_c > 0$ such that

$$\sup_{\mathbf{x}_0 \in \mathbb{R}^n} T(\mathbf{x}_0) \leq T_c, \quad (2)$$

where $T(\mathbf{x}_0)$ is the physical time required for the state starting from $\mathbf{x}_0$ to reach the origin, then the origin is globally predefined-time stable.

Lemma 1 ([31]). For the above nonlinear system, suppose that there exists a continuously differentiable, positive-definite, and radially unbounded Lyapunov function $V : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$. If for a given time constant $T_c > 0$ and a parameter $\eta \in (0, 1)$, the derivative of V along the system trajectory satisfies

$$\dot{V} \leq -\frac{\pi}{\eta T_c} \left(V(\mathbf{x}(t))^{1-\eta/2} + V(\mathbf{x}(t))^{1+\eta/2} \right) + b, \quad (3)$$

then the origin is globally predefined-time stable, the convergence-time upper bound is determined by the physical constant $2T_c$, and V converges into the region $V \leq \frac{\eta b T_c}{\pi}$.

Lemma 2 ([22]). For any real sequence $\{y_j\}_{j=1}^m \in \mathbb{R}$ and any $\eta \in (0, 1)$, the following inequalities hold:

$$\left(\sum_{j=1}^m y_j^2 \right)^{\frac{2-\eta}{2}} \leq \sum_{j=1}^m |y_j|^{2-\eta}, \quad (4)$$

$$m^{-\eta/2} \left(\sum_{j=1}^m y_j^2 \right)^{\frac{2+\eta}{2}} \leq \sum_{j=1}^m |y_j|^{2+\eta}. \quad (5)$$

2.2. Dynamics of Highly Flexible Aircraft

Following the control-oriented HFA model used by Peng and Ma [2], the aircraft is composed of three identical rigid wing bodies connected through hinges. Each rigid panel includes a propeller for thrust production, an aileron along the aft of the main wing, and an elevator attached at the end of the boom [33]. Because pronounced structural deformation is concentrated at the hinged interfaces between adjacent wing panels, the model can still

capture the response characteristics associated with flexible structural deflections. The aircraft under consideration is shown in Figure 1.

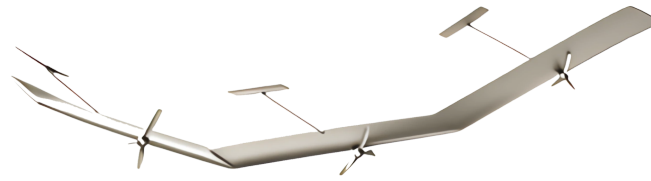


Figure 1. The three rigid model for a highly flexible aircraft.

The longitudinal control-oriented nonlinear model is written as

$$\begin{cases} \dot{V} = \frac{T \cos \alpha - D}{m} - g \sin \gamma, \\ \dot{\alpha} = -\frac{T \sin \alpha + L}{mV} + q + \frac{g \cos \gamma}{V}, \\ \dot{h} = V \sin \gamma, \\ \dot{\theta} = q, \\ \dot{q} = \frac{M - 2c_2 \sin(\eta) \cos(\eta) \dot{\eta} q}{c_1 + c_2 \sin^2(\eta)}, \\ \dot{\eta} = \frac{H - \kappa_c \dot{\eta} - \kappa_k \eta + d_1 - d_2}{d_3}, \end{cases} \quad (6)$$

where V is velocity, $\gamma = \theta - \alpha$ is the flight-path angle, h is altitude, α is the angle of attack, θ is pitch angle, q is pitch rate, η is the elastic dihedral coordinate, T is total thrust, D is drag, L is lift, and M and H are the total moment and angular moment related to control inputs. The coefficients are

$$\begin{aligned} c_1 &= 3I_{yy}^*, & c_2 &= 2I_{zz}^* - 2I_{yy}^* + \frac{m^* s^2}{6}, \\ d_1 &= \frac{s}{2} m^* \left[(\dot{V} \sin \alpha + V \cos \alpha \dot{\alpha}) \cos \eta - V \sin \alpha \sin \eta \dot{\eta} - \frac{2s}{3} \cos \eta \sin \eta \dot{\eta}^2 \right], \\ d_2 &= \left(I_{yy}^* - I_{zz}^* - \frac{m^* s^2}{12} \right) \sin \eta \cos \eta q^2 - \frac{m^* s}{2} \cos \eta V \cos \alpha q, \\ d_3 &= I_{x_3 x_3}^* + m^* \left(\frac{s^2}{4} + \frac{s^2}{6} \cos^2 \eta \right). \end{aligned} \quad (7)$$

The compact nonlinear HFA model is

$$\dot{\mathbf{X}} = \mathbf{f}(\mathbf{X}, \mathbf{U}), \quad \mathbf{X} = [V, \alpha, h, \theta, q, \eta, \dot{\eta}]^T, \quad (8)$$

with the five-input vector

$$\mathbf{U} = [\delta_t, \delta_{a,c}, \delta_{a,o}, \delta_{e,c}, \delta_{e,o}]^T. \quad (9)$$

After linearization around a trim point $(\mathbf{X}_0, \mathbf{U}_0)$ satisfying $\mathbf{f}(\mathbf{X}_0, \mathbf{U}_0) = \mathbf{0}$, the local model is

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}, \quad (10)$$

where $\mathbf{x} = \mathbf{X} - \mathbf{X}_0$ and $\mathbf{u} = \mathbf{U} - \mathbf{U}_0$. In the simulation of this paper, the altitude deviation is not controlled explicitly, and the six-state vector is

$$\mathbf{x}(t) = [V(t), \alpha(t), \theta(t), q(t), \eta(t), \dot{\eta}(t)]^T \in \mathbb{R}^6, \quad (11)$$

with the five-input vector $\mathbf{u}(t) \in \mathbb{R}^5$ inherited from $\mathbf{U}$.

2.3. Disturbed Tracking Error Dynamics and Control Objective

Consider the linearized HFA system subject to unknown external disturbance:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{d}(t), \quad (12)$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is the system state vector, $\mathbf{u}(t) \in \mathbb{R}^m$ is the control input vector, $\mathbf{d}(t) \in \mathbb{R}^n$ is the unknown external bounded disturbance, $\mathbf{A} \in \mathbb{R}^{n \times n}$ is the system matrix, and $\mathbf{B} \in \mathbb{R}^{n \times m}$ is the input coupling matrix.

For tracking control, let the desired reference signal be $\mathbf{x}_d(t)$. The error system is

$$\dot{\mathbf{z}}(t) = \mathbf{F}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{d}(t), \quad (13)$$

where

$$\mathbf{z}(t) = \mathbf{x}(t) - \mathbf{x}_d(t), \quad \mathbf{F}(t) = \mathbf{A}\mathbf{x}(t) - \dot{\mathbf{x}}_d(t), \quad (14)$$

and $\dot{\mathbf{x}}_d(t)$ denotes the derivative of the reference signal. In this paper, $\|\cdot\|$ denotes the vector 2-norm or the induced matrix 2-norm.

The control objective is to design a zero-sum reinforcement learning controller, a predefined-time prescribed-performance mechanism, and a predefined-time critic update law such that the following requirements are achieved: (i) the selected tracking-error components satisfy prescribed transient and steady-state accuracy envelopes; (ii) the neural-network parameter estimation error $\tilde{\mathbf{W}}(t)$ converges into a bounded neighborhood within the strict user-defined time T_w ; and (iii) under the external disturbance $\mathbf{d}(\mathbf{z}(t))$, the coupled physical-learning closed-loop system remains uniformly ultimately bounded.

3. Zero-Sum Game-Based Predefined-Time Prescribed-Performance RL Controller Design

3.1. Predefined-Time Prescribed-Performance Constraints and Barrier Functions

For each constrained tracking-error channel $z_i(t)$, define the predefined-time prescribed-performance function

$$\sigma_i(t) = (\sigma_{i,0} - \sigma_{i,\infty})q(t) + \sigma_{i,\infty}, \quad (15)$$

where $\sigma_{i,0} > \sigma_{i,\infty} > 0$, and

$$q(t) = \begin{cases} \exp\left(\frac{\alpha_\sigma t}{t - T_p}\right), & 0 \leq t < T_p, \\ 0, & t \geq T_p. \end{cases} \quad (16)$$

Here, $T_p > 0$ is the prescribed performance-settling time and $\alpha_\sigma > 0$ is a shaping parameter. Since $q(0) = 1$ and $q(t) = 0$ for $t \geq T_p$, $\sigma_i(t)$ decreases from $\sigma_{i,0}$ to $\sigma_{i,\infty}$ within the prescribed time T_p . The predefined-time prescribed-performance constraint is

$$-\sigma_i(t) < z_i(t) < \sigma_i(t), \quad i \in \mathcal{Z}_p, \quad (17)$$

where $\mathcal{Z}_p$ is the set of error channels constrained by the performance envelopes.

To transform the bilateral inequality (17) into positivity constraints, define the upper and lower boundary-margin functions as

$$\phi_{i,U}(t, z_i) = \sigma_i(t) - z_i(t), \quad (18)$$

$$\phi_{i,L}(t, z_i) = \sigma_i(t) + z_i(t). \quad (19)$$

Then (17) is equivalent to

$$\phi_{i,U}(t, z_i) > 0, \quad \phi_{i,L}(t, z_i) > 0. \quad (20)$$

Therefore, keeping the two boundaries positive during the closed-loop evolution directly guarantees that $z_i(t)$ remains inside the predefined-time prescribed-performance envelope.

Let $\mathcal{D}_p$ denote the directly actuated constrained channels satisfying $B_{i,:} \neq \mathbf{0}$. For $i \in \mathcal{D}_p$, the control input can directly affect $\dot{z}_i$. The logarithmic prescribed-performance barrier is defined by

$$\mathcal{B}_i^d(\mathbf{z}, t) = k_{d,i} \left[\ln \left(\frac{\phi_{i,U}(t, z_i)}{\phi_{i,L}(t, z_i)} \right) \right]^2, \quad k_{d,i} > 0, \quad (21)$$

or, after substituting (18) and (19),

$$\mathcal{B}_i^d(\mathbf{z}, t) = k_{d,i} \left[\ln \left(\frac{\sigma_i(t) - z_i(t)}{\sigma_i(t) + z_i(t)} \right) \right]^2. \quad (22)$$

When $z_i(t)$ approaches either boundary, $\phi_{i,U}$ or $\phi_{i,L}$ approaches zero and the barrier tends to infinity, which creates a boundary-repulsion effect in the optimal feedback law.

For the channels on which the control input cannot directly act in the system structure, taking the third and fifth rows of the input matrix as zero, namely $B_{3,:} = \mathbf{0}$ and $B_{5,:} = \mathbf{0}$, as an example, $\mathbf{u}$ cannot directly appear in $\dot{z}_3$ and $\dot{z}_5$. From the kinematic structure, one has

$$\dot{z}_3(t) = z_4(t) + \Delta_3(t), \quad (23)$$

$$\dot{z}_5(t) = z_6(t) + \Delta_5(t), \quad (24)$$

where $\Delta_3(t) = x_{d,4}(t) - \dot{x}_{d,3}(t)$ and $\Delta_5(t) = x_{d,6}(t) - \dot{x}_{d,5}(t)$ collect the reference-induced mismatch terms. To express the construction uniformly, define the high-relative-degree error-index set and the corresponding derivative-channel map as

$$\mathcal{H} = \{3, 5\}, \quad \nu(3) = 4, \quad \nu(5) = 6. \quad (25)$$

Then, for any $i \in \mathcal{H}$,

$$\dot{z}_i(t) = z_{\nu(i)}(t) + \Delta_i(t). \quad (26)$$

The predefined-time prescribed performance is still imposed on the original error $z_i(t)$, namely,

$$-\sigma_i(t) < z_i(t) < \sigma_i(t), \quad i \in \mathcal{H}_p \subseteq \mathcal{H}. \quad (27)$$

To handle the relative-degree mismatch caused by $B_{i,:} = \mathbf{0}$, introduce the first-order recursive boundary

$$r_{i,U}^\circ = \dot{\phi}_{i,U} + \lambda_i \phi_{i,U}, \quad \lambda_i > 0, \quad (28)$$

$$r_{i,L}^\circ = \dot{\phi}_{i,L} + \lambda_i \phi_{i,L}. \quad (29)$$

Suppose that the uncertain mismatch satisfies $|\Delta_i(t)| \leq \bar{\Delta}_i$. From (18), (19), and (26), the robust lower-bound recursive boundary used for control design are

$$r_{i,U}(z, t) = \dot{\sigma}_i(t) - z_{v(i)}(t) - \bar{\Delta}_i + \lambda_i(\sigma_i(t) - z_i(t)), \quad (30)$$

$$r_{i,L}(z, t) = \dot{\sigma}_i(t) + z_{v(i)}(t) - \bar{\Delta}_i + \lambda_i(\sigma_i(t) + z_i(t)). \quad (31)$$

If $r_{i,U} > 0$ and $r_{i,L} > 0$, then $r_{i,U}^\circ > 0$ and $r_{i,L}^\circ > 0$ are guaranteed under the bounded uncertainty. Hence,

$$\dot{\phi}_{i,U} + \lambda_i \phi_{i,U} > 0, \quad \dot{\phi}_{i,L} + \lambda_i \phi_{i,L} > 0. \quad (32)$$

Equivalently,

$$\frac{d}{dt} \left[e^{\lambda_i t} \phi_{i,U}(t, z_i(t)) \right] > 0. \quad (33)$$

Therefore, if $\phi_{i,U}(0, z_i(0)) > 0$, then

$$\phi_{i,U}(t, z_i(t)) > e^{-\lambda_i t} \phi_{i,U}(0, z_i(0)) > 0. \quad (34)$$

Similarly, if $\phi_{i,L}(0, z_i(0)) > 0$, then

$$\phi_{i,L}(t, z_i(t)) > e^{-\lambda_i t} \phi_{i,L}(0, z_i(0)) > 0. \quad (35)$$

Thus, $r_{i,U}$ and $r_{i,L}$ are the high-relative-degree recursive boundary that prevent the original error $z_i(t)$ from crossing the predefined-time prescribed-performance boundary.

For $i \in \mathcal{H}_p$, construct the high-relative-degree prescribed-performance barrier as

$$\mathcal{B}_i^h(z, t) = k_{0,i} \left[\ln \left(\frac{\phi_{i,U}}{\phi_{i,L}} \right) \right]^2 + k_{1,i} \left[\ln \left(\frac{r_{i,U}}{r_{i,L}} \right) \right]^2, \quad i \in \mathcal{H}_p, \quad (36)$$

where $k_{0,i} > 0$ and $k_{1,i} > 0$ are barrier weights. The first term directly corresponds to the prescribed boundary of $z_i(t)$. The second term corresponds to the recursive-margin positivity constraint and transmits the boundary-keeping requirement to the channel $z_{v(i)}(t)$ through which the control input can act.

Putting the directly actuated error channels and high-relative-degree error channels together, the predefined-time prescribed-performance barrier function is defined as

$$\mathcal{B}_p(z, t) = \sum_{i \in \mathcal{D}_p} \mathcal{B}_i^d(z, t) + \sum_{i \in \mathcal{H}_p} \mathcal{B}_i^h(z, t). \quad (37)$$

Here, $\mathcal{D}_p$ denotes the directly actuated performance-constraint channel set, and $\mathcal{H}_p$ denotes the high-relative-degree performance-constraint channel set. If predefined-time prescribed-performance constraints are imposed only on z_1, z_2 , and z_5 , one may take $\mathcal{D}_p = \{1, 2\}$ and $\mathcal{H}_p = \{5\}$. If constraints are imposed simultaneously on z_3 and z_5 in addition to z_1 and z_2 , one may take $\mathcal{D}_p = \{1, 2\}$ and $\mathcal{H}_p = \{3, 5\}$.

The corresponding gradients are required by the HJI feedback law and the Bellman residual. For a directly actuated channel, differentiating (22) with respect to z gives

$$\nabla_z \mathcal{B}_i^d = -\frac{4k_{d,i}\sigma_i(t)}{\sigma_i^2(t) - z_i^2(t)} \ln \left(\frac{\sigma_i(t) - z_i(t)}{\sigma_i(t) + z_i(t)} \right) e_i, \quad (38)$$

where $e_i \in \mathbb{R}^n$ is the i th unit vector. For the first term of the high-relative-degree barrier,

$$\nabla_z \mathcal{B}_{i,0}^h = -\frac{4k_{0,i}\sigma_i(t)}{\sigma_i^2(t) - z_i^2(t)} \ln \left(\frac{\sigma_i(t) - z_i(t)}{\sigma_i(t) + z_i(t)} \right) e_i. \quad (39)$$

For the recursive-margin term,

$$\nabla_z \mathcal{B}_{i,1}^h = 2k_{1,i} \ln\left(\frac{r_{i,U}}{r_{i,L}}\right) \left(\frac{\nabla_z r_{i,U}}{r_{i,U}} - \frac{\nabla_z r_{i,L}}{r_{i,L}} \right), \quad (40)$$

where

$$\nabla_z r_{i,U} = -\mathbf{e}_{v(i)} - \lambda_i \mathbf{e}_i, \quad \nabla_z r_{i,L} = \mathbf{e}_{v(i)} + \lambda_i \mathbf{e}_i. \quad (41)$$

Using (30), (31) and (40) becomes

$$\nabla_z \mathcal{B}_{i,1}^h = -\frac{4k_{1,i}(\dot{\sigma}_i(t) + \lambda_i \sigma_i(t) - \bar{\Delta}_i)}{r_{i,U} r_{i,L}} \ln\left(\frac{r_{i,U}}{r_{i,L}}\right) (\mathbf{e}_{v(i)} + \lambda_i \mathbf{e}_i). \quad (42)$$

Consequently, the total high-relative-degree barrier gradient is

$$\begin{aligned} \nabla_z \mathcal{B}_i^h = & -\frac{4k_{0,i}\sigma_i(t)}{\sigma_i^2(t) - z_i^2(t)} \ln\left(\frac{\sigma_i(t) - z_i(t)}{\sigma_i(t) + z_i(t)}\right) \mathbf{e}_i \\ & - \frac{4k_{1,i}(\dot{\sigma}_i(t) + \lambda_i \sigma_i(t) - \bar{\Delta}_i)}{r_{i,U} r_{i,L}} \ln\left(\frac{r_{i,U}}{r_{i,L}}\right) (\mathbf{e}_{v(i)} + \lambda_i \mathbf{e}_i), \quad i \in \mathcal{H}_p. \end{aligned} \quad (43)$$

These expressions give the actual implementation of predefined-time prescribed performance in the proposed RL controller: T_p fixes when the envelope reaches $\sigma_{i,\infty}$, (20) and (32) state the required positive-margin conditions, (21)–(36) convert these conditions into barrier penalties, and (38)–(43) inject the corresponding boundary-repulsion directions into the HJI policy and the Bellman residual.

3.2. Zero-Sum Game Optimal Predefined-Time Prescribed-Performance Control

For the error system (13), define the nonlinear utility function $\mathcal{J} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}$ as

$$\mathcal{J}(\mathbf{z}(t), \mathbf{u}(t), \mathbf{d}(t)) = \mathbf{z}^T(t) \mathbf{Q} \mathbf{z}(t) + \mathbf{u}^T(t) \mathbf{R} \mathbf{u}(t) + \mathcal{B}_p(\mathbf{z}(t), t) - \delta^2 \mathbf{d}^T(t) \mathbf{d}(t), \quad (44)$$

where $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is the tracking-error weight, $\mathbf{R} \in \mathbb{R}^{m \times m}$ is the control-input weight, and $\delta > 0$ is the disturbance attenuation parameter. The barrier term penalizes approaching the predefined-time prescribed-performance boundary, whereas the negative disturbance-energy term reflects the adversarial role of the disturbance in the zero-sum game.

The optimal value function $V_c(t)$ is defined as the min-max functional over the infinite horizon:

$$V_c(t) = \min_u \max_d \int_0^\infty \mathcal{J}(\mathbf{z}(t), \mathbf{u}(t), \mathbf{d}(t)) dt. \quad (45)$$

For $V_c(t)$, the differential-game Hamiltonian is

$$\begin{aligned} H = & \mathbf{z}^T(t) \mathbf{Q} \mathbf{z}(t) + \mathbf{u}^T(t) \mathbf{R} \mathbf{u}(t) + \mathcal{B}_p(\mathbf{z}(t), t) - \delta^2 \mathbf{d}^T(t) \mathbf{d}(t) \\ & + \nabla V_c^T(\mathbf{z}(t)) (\mathbf{F}(t) + \mathbf{B} \mathbf{u}(t) + \mathbf{d}(t)). \end{aligned} \quad (46)$$

Taking the extremum derivatives with respect to the control and disturbance variables gives

$$\frac{\partial H}{\partial \mathbf{u}(t)} = 2\mathbf{R} \mathbf{u}(t) + \mathbf{B}^T \nabla V_c(t) = \mathbf{0}, \quad (47)$$

$$\frac{\partial H}{\partial \mathbf{d}(t)} = -2\delta^2 \mathbf{d}(t) + \nabla V_c(\mathbf{z}(t)) = \mathbf{0}. \quad (48)$$

Therefore, the optimal feedback Nash strategies are

$$\mathbf{u}^*(t) = -\frac{1}{2} \mathbf{R}^{-1} \mathbf{B}^T \nabla V_c^*(t), \quad \mathbf{d}^*(t) = \frac{1}{2\delta^2} \nabla V_c^*(t). \quad (49)$$

According to the extremum conditions, the optimal feedback strategies satisfy the steady-state Hamilton–Jacobi–Isaacs equation

$$\frac{\partial V_c^*(z, t)}{\partial t} + H(z, u^*(z, t), d^*(z, t), t) = 0 \quad (50)$$

3.3. Critic Neural-Network Approximation

A critic neural network is adopted to parameterize and reconstruct the unknown smooth part of the optimal value function. Since the predefined-time prescribed-performance barrier is known, it is explicitly embedded into the value-function approximation as

$$V_c^*(z(t), t) = W^{*T} \psi(z(t)) + \mathcal{B}_p(z(t), t) + \varepsilon(z(t)), \quad (51)$$

where $W^* \in \mathbb{R}^M$ is the ideal network weight vector, M is the number of hidden nodes, and $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^M$ is the activation basis vector. Taking the gradient with respect to the state yields

$$\nabla V_c^*(z(t), t) = \nabla \psi^T(z(t)) W^* + \nabla_z \mathcal{B}_p(z(t), t) + \nabla \varepsilon(z(t)). \quad (52)$$

The approximation residual satisfies

$$\|\varepsilon(z(t))\| \leq \varepsilon_c, \quad \|\nabla \varepsilon(z(t))\| \leq \varepsilon_{dc} \quad (53)$$

Substituting (52) into (49) gives

$$u^*(t) = -\frac{1}{2} R^{-1} B^T \left(\nabla \psi^T(z(t)) W^* + \nabla_z \mathcal{B}_p(z(t), t) + \nabla \varepsilon(z(t)) \right), \quad (54)$$

$$d^*(t) = \frac{1}{2\delta^2} \left(\nabla \psi^T(z(t)) W^* + \nabla_z \mathcal{B}_p(z(t), t) + \nabla \varepsilon(z(t)) \right). \quad (55)$$

Using the online estimated weight $\hat{W}(t) \in \mathbb{R}^M$, the implementable control and disturbance strategies are constructed as

$$\begin{aligned} \hat{u}(z(t)) &= -\frac{1}{2} R^{-1} B^T \left(\nabla \psi^T(z(t)) \hat{W}(t) + \nabla_z \mathcal{B}_p(z(t), t) \right), \\ \hat{d}(z(t)) &= \frac{1}{2\delta^2} \left(\nabla \psi^T(z(t)) \hat{W}(t) + \nabla_z \mathcal{B}_p(z(t), t) \right). \end{aligned} \quad (56)$$

Define $\sigma(t) \in \mathbb{R}^M$ as

$$\sigma(t) = \nabla \psi(z(t)) \left(F(t) + B \hat{u}(t) + \hat{d}(z(t)) \right). \quad (57)$$

The Bellman residual $e(t) \in \mathbb{R}$ is

$$\begin{aligned} e(t) &= \frac{\partial \mathcal{B}_p(z(t), t)}{\partial t} + \hat{W}^T(t) \sigma(t) + \nabla_z \mathcal{B}_p^T(z(t), t) \left(F(t) + B \hat{u}(t) + \hat{d}(z(t)) \right) \\ &\quad + z^T(t) Q z(t) + \hat{u}^T(t) R \hat{u}(t) + \mathcal{B}_p(z(t), t) - \delta^2 \hat{d}^T(t) \hat{d}(t). \end{aligned} \quad (58)$$

where $\frac{\partial \mathcal{B}_p}{\partial t} = \sum_{i \in \mathcal{D}_p} \frac{\partial \mathcal{B}_{p,i}^d}{\partial \sigma_i} \dot{\sigma}_i + \sum_{i \in \mathcal{H}_p} \left(\frac{\partial \mathcal{B}_{p,i}^h}{\partial \sigma_i} \dot{\sigma}_i + \frac{\partial \mathcal{B}_{p,i}^h}{\partial \tilde{\sigma}_i} \dot{\tilde{\sigma}}_i \right)$. Define the online parameter-estimation error as

$$\widetilde{W}(t) = \hat{W}(t) - W^*. \quad (59)$$

By algebraic subtraction from the ideal HJI equation, the residual equation is converted into the explicit mapping

$$e(t) = \widetilde{\mathbf{W}}^T(t)\boldsymbol{\sigma}(t) + \varepsilon_H(t), \quad (60)$$

where $\varepsilon_H(t)$ represents the residual induced by the limited approximation ability of the critic network, and the approximation of the worst disturbance. On a compact set $\mathcal{Z}$, it satisfies

$$\sup_{z(t) \in \mathcal{Z}} |\varepsilon_H(t)| \leq \varepsilon_{Hd}. \quad (61)$$

Adaptive control and disturbance-observer-based control can provide efficient online uncertainty compensation, and high-order sliding-mode control offers strong robustness with relatively direct implementation. In this work, the disturbance is represented as an adversarial player in a min–max learning problem, the transient tracking requirements are embedded through prescribed-performance barriers, and the critic learning error is assigned a practical predefined-time convergence bound. Therefore, the proposed framework is mainly attractive when worst-case game interpretation, prescribed transient performance, and data-driven policy learning are required simultaneously.

3.4. Predefined-Time Adaptive Critic Update Law Design

To make the sufficient-excitation assumption valid in the learning law, the critic update is constructed by using both the current Hamiltonian estimation error and a finite stack of recorded data. Let

$$\mathcal{D}_h(t) = \left\{ t_j, z(t_j), \hat{\mathbf{u}}(t_j), \hat{\mathbf{d}}(t_j), \boldsymbol{\sigma}(t_j) \right\}_{j=1}^{N_h} \quad (62)$$

denote the recorded data set collected before the current time t , where $N_h \geq M$. For each stored sample t_j , the historical Bellman residual is evaluated by the current critic weight $\hat{\mathbf{W}}(t)$ as

$$\begin{aligned} e(t_j, t) = & \frac{\partial \mathcal{B}_p(z, \tau)}{\partial \tau} \Big|_{z=z(t_j), \tau=t_j} + \hat{\mathbf{W}}^T(t)\boldsymbol{\sigma}(t_j) + \nabla_z \mathcal{B}_p^T(z(t_j), t_j) \left(\mathbf{F}(t_j) + \mathbf{B}\hat{\mathbf{u}}(t_j) + \hat{\mathbf{d}}(t_j) \right) \\ & + \mathbf{z}^T(t_j)\mathbf{Q}\mathbf{z}(t_j) + \hat{\mathbf{u}}^T(t_j)\mathbf{R}\hat{\mathbf{u}}(t_j) + \mathcal{B}_p(z(t_j), t_j) - \delta^2 \hat{\mathbf{d}}^T(t_j)\hat{\mathbf{d}}(t_j), \\ & j = 1, 2, \dots, N_h. \end{aligned} \quad (63)$$

That is, the state, policy, regressor, and barrier-gradient information at t_j is stored, while the unknown ideal parameter is still replaced by the current online estimate $\hat{\mathbf{W}}(t)$. Define the normalized current and recorded regressors as

$$\begin{aligned} \bar{\boldsymbol{\sigma}}(t) &= \frac{\boldsymbol{\sigma}(t)}{1 + \boldsymbol{\sigma}^T(t)\boldsymbol{\sigma}(t)}, & \bar{e}(t) &= \frac{e(t)}{1 + \boldsymbol{\sigma}^T(t)\boldsymbol{\sigma}(t)}, \\ \bar{\boldsymbol{\sigma}}(t_j) &= \frac{\boldsymbol{\sigma}(t_j)}{1 + \boldsymbol{\sigma}^T(t_j)\boldsymbol{\sigma}(t_j)}, & \bar{e}(t_j, t) &= \frac{e(t_j, t)}{1 + \boldsymbol{\sigma}^T(t_j)\boldsymbol{\sigma}(t_j)}. \end{aligned} \quad (64)$$

Then the current and historical normalized Bellman residuals can be written as

$$\bar{e}(t) = \bar{\boldsymbol{\sigma}}^T(t)\widetilde{\mathbf{W}}(t) + \bar{\varepsilon}_H(t), \quad (65)$$

$$\bar{e}(t_j, t) = \bar{\boldsymbol{\sigma}}^T(t_j)\widetilde{\mathbf{W}}(t) + \bar{\varepsilon}_H(t_j), \quad j = 1, 2, \dots, N_h, \quad (66)$$

where $\bar{\varepsilon}_H(t)$ and $\bar{\varepsilon}_H(t_j)$ are the normalized approximation residuals. On the compact set $\mathcal{Z}$, there exists a positive constant $\bar{\varepsilon}_{Hd}$ such that

$$\max\{|\bar{\varepsilon}_H(t)|, |\bar{\varepsilon}_H(t_1)|, \dots, |\bar{\varepsilon}_H(t_{N_h})|\} \leq \bar{\varepsilon}_{Hd}. \quad (67)$$

Let the recorded-data matrix be

$$\mathbf{\Pi} = [\bar{\sigma}(t_1), \bar{\sigma}(t_2), \dots, \bar{\sigma}(t_{N_h})] \in \mathbb{R}^{M \times N_h}, \quad (68)$$

and define

$$\mathbf{\Sigma} = \mathbf{\Pi} \mathbf{\Pi}^T = \sum_{j=1}^{N_h} \bar{\sigma}(t_j) \bar{\sigma}^T(t_j). \quad (69)$$

Assumption 1. The recorded historical data set satisfies the sufficient excitation condition. Namely, the data stack $\{\bar{\sigma}(t_j)\}_{j=1}^{N_h}$ is N_h -sufficiently rich and

$$\text{rank}(\mathbf{\Pi}) = M, \quad (70)$$

or equivalently, there exists a constant $\lambda_0 > 0$ such that

$$\lambda_{\min}(\mathbf{\Sigma}) \geq \lambda_0 > 0. \quad (71)$$

Remark 1. Assumption 1 is a finite-data sufficient-excitation condition rather than a persistent-excitation condition. Before the predefined-time online learning interval begins, candidate samples are collected until $\text{rank}(\mathbf{\Pi}) = M$ is verified. During online execution, the historical stack can be maintained by retaining or replacing samples so that this rank condition is preserved. New candidate samples can replace existing samples only when the minimum singular value of the stored regressor matrix is not reduced. If the full-rank condition cannot be satisfied, the practical predefined-time convergence result of Theorem 1 cannot be guaranteed. The critic update is obtained from the fractional-power minimization of both current and historical Bellman residuals. Specifically, define

$$\begin{aligned} E(\hat{\mathbf{W}}(t)) &= \frac{1}{2-\eta} \left(|\bar{e}(t)|^{2-\eta} + \sum_{j=1}^{N_h} |\bar{e}(t_j, t)|^{2-\eta} \right) \\ &\quad + \frac{1}{2+\eta} \left(|\bar{e}(t)|^{2+\eta} + \sum_{j=1}^{N_h} |\bar{e}(t_j, t)|^{2+\eta} \right), \end{aligned} \quad (72)$$

where $\eta \in (0, 1)$. To guarantee that the learning convergence time is strictly constrained within T_w , the following current-and-history network parameter update law is designed:

$$\begin{aligned} \dot{\hat{\mathbf{W}}}(t) &= -(\alpha_1 + 1) \mathbf{\Gamma} \bar{\sigma}(t) \text{sig}^{1-\eta}(\bar{e}(t)) - (\alpha_1 + 1) \mathbf{\Gamma} \sum_{j=1}^{N_h} \bar{\sigma}(t_j) \text{sig}^{1-\eta}(\bar{e}(t_j, t)) \\ &\quad - (\alpha_2 + 1) \mathbf{\Gamma} \bar{\sigma}(t) \text{sig}^{1+\eta}(\bar{e}(t)) - (\alpha_2 + 1) \mathbf{\Gamma} \sum_{j=1}^{N_h} \bar{\sigma}(t_j) \text{sig}^{1+\eta}(\bar{e}(t_j, t)). \end{aligned} \quad (73)$$

The sign-power operator is defined by

$$\text{sig}^p(x) = |x|^p \text{sgn}(x). \quad (74)$$

The gain matrix $\mathbf{\Gamma} \in \mathbb{R}^{M \times M}$ is positive-definite, symmetric, and diagonal, satisfying $\mathbf{\Gamma} = \mathbf{\Gamma}^T > 0$. Introduce an analytical constant $\theta \in (0, 1)$ and define

$$\begin{aligned} c_1 &= (2\lambda_{\min}(\mathbf{\Gamma})\lambda_0)^{1-\eta/2}, \\ c_2 &= N_h^{-\eta/2} (2\lambda_{\min}(\mathbf{\Gamma})\lambda_0)^{1+\eta/2}. \end{aligned} \quad (75)$$

The learning factors α_1 and α_2 are selected as

$$\alpha_1 = \frac{\pi}{\theta T'_w \eta c_1}, \quad \alpha_2 = \frac{\pi}{\theta T'_w \eta c_2}, \quad (76)$$

where $T'_w = 0.5T_w$ is the auxiliary learning-time constant, so the predefined upper bound for the critic-weight learning process is $T_w = 2T'_w$.

The structure diagram of the algorithm in this article is shown in Figure 2.

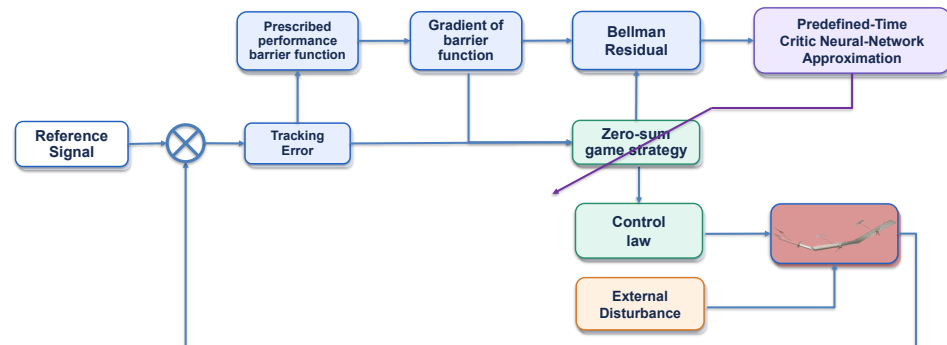


Figure 2. Structure diagram of the proposed algorithm.

Although the proposed method employs a critic neural network, it does not involve online multilayer-network training or backpropagation. The critic basis functions and their analytical gradients are prescribed a priori, and only the M -dimensional critic output-weight vector is updated online. Let n , m , M , and N_h denote the state dimension, control dimension, number of critic basis functions, and number of stored samples, respectively. The current policy and Bellman-residual evaluations require $O(Mn + nm + n^2 + m^2)$ operations, while the current-and-history critic update requires $O(M(N_h + 1))$ operations because the sample-dependent quantities are fixed after storage. Hence, the total online complexity is $O(Mn + MN_h + nm + n^2 + m^2)$, and the persistent memory requirement is $O(N_h(M + n + m) + M)$. The implementation therefore involves only finite matrix–vector operations and historical-data summations, without online matrix inversion, backpropagation, or iterative optimization, supporting real-time implementation for moderate M and N_h .

4. Stability Analysis

Theorem 1. Under Assumption 1, if the current-and-history fractional-power network update law (73) is applied, the practical predefined-time stability of the critic approximation error can be guaranteed.

Proof. Construct the Lyapunov functional reflecting the critic error energy:

$$V_w(\tilde{\mathbf{W}}(t)) = \frac{1}{2} \tilde{\mathbf{W}}^T(t) \Gamma^{-1} \tilde{\mathbf{W}}(t). \quad (77)$$

Taking its derivative and substituting the current-and-history update law (73) gives

$$\begin{aligned} \dot{V}_w(t) = & -(\alpha_1 + 1) \tilde{\sigma}^T(t) \tilde{\mathbf{W}}(t) \operatorname{sig}^{1-\eta}(\tilde{e}(t)) - (\alpha_1 + 1) \sum_{j=1}^{N_h} \tilde{\sigma}^T(t_j) \tilde{\mathbf{W}}(t) \operatorname{sig}^{1-\eta}(\tilde{e}(t_j, t)) \\ & - (\alpha_2 + 1) \tilde{\sigma}^T(t) \tilde{\mathbf{W}}(t) \operatorname{sig}^{1+\eta}(\tilde{e}(t)) - (\alpha_2 + 1) \sum_{j=1}^{N_h} \tilde{\sigma}^T(t_j) \tilde{\mathbf{W}}(t) \operatorname{sig}^{1+\eta}(\tilde{e}(t_j, t)). \end{aligned} \quad (78)$$

Using (65) and (66), Equation (78) can be rewritten as

$$\begin{aligned}\dot{V}_w(t) = & -(\alpha_1 + 1)\bar{\sigma}^T(t)\widetilde{\mathbf{W}}(t)\text{sig}^{1-\eta}(\bar{\sigma}^T(t)\widetilde{\mathbf{W}}(t) + \bar{\varepsilon}_H(t)) \\ & -(\alpha_1 + 1)\sum_{j=1}^{N_h}\bar{\sigma}^T(t_j)\widetilde{\mathbf{W}}(t)\text{sig}^{1-\eta}(\bar{\sigma}^T(t_j)\widetilde{\mathbf{W}}(t) + \bar{\varepsilon}_H(t_j)) \\ & -(\alpha_2 + 1)\bar{\sigma}^T(t)\widetilde{\mathbf{W}}(t)\text{sig}^{1+\eta}(\bar{\sigma}^T(t)\widetilde{\mathbf{W}}(t) + \bar{\varepsilon}_H(t)) \\ & -(\alpha_2 + 1)\sum_{j=1}^{N_h}\bar{\sigma}^T(t_j)\widetilde{\mathbf{W}}(t)\text{sig}^{1+\eta}(\bar{\sigma}^T(t_j)\widetilde{\mathbf{W}}(t) + \bar{\varepsilon}_H(t_j)).\end{aligned}\quad (79)$$

For the ideal approximation case, namely $\bar{\varepsilon}_H(t) = 0$ and $\bar{\varepsilon}_H(t_j) = 0, j = 1, \dots, N_h$, one obtains

$$\begin{aligned}\dot{V}_w(t) = & -(\alpha_1 + 1)\left|\bar{\sigma}^T(t)\widetilde{\mathbf{W}}(t)\right|^{2-\eta} - (\alpha_2 + 1)\left|\bar{\sigma}^T(t)\widetilde{\mathbf{W}}(t)\right|^{2+\eta} \\ & -(\alpha_1 + 1)\sum_{j=1}^{N_h}\left|\bar{\sigma}^T(t_j)\widetilde{\mathbf{W}}(t)\right|^{2-\eta} - (\alpha_2 + 1)\sum_{j=1}^{N_h}\left|\bar{\sigma}^T(t_j)\widetilde{\mathbf{W}}(t)\right|^{2+\eta}.\end{aligned}\quad (80)$$

Since the current term is non-positive, it can be retained in the learning law but discarded in the upper-bound estimate. Hence,

$$\dot{V}_w(t) \leq -\alpha_1\sum_{j=1}^{N_h}\left|\bar{\sigma}^T(t_j)\widetilde{\mathbf{W}}(t)\right|^{2-\eta} - \alpha_2\sum_{j=1}^{N_h}\left|\bar{\sigma}^T(t_j)\widetilde{\mathbf{W}}(t)\right|^{2+\eta}.\quad (81)$$

According to Lemma 2, one has

$$\sum_{j=1}^{N_h}\left|\bar{\sigma}^T(t_j)\widetilde{\mathbf{W}}(t)\right|^{2-\eta} \geq \left(\sum_{j=1}^{N_h}\left|\bar{\sigma}^T(t_j)\widetilde{\mathbf{W}}(t)\right|^2\right)^{1-\eta/2},\quad (82)$$

$$\sum_{j=1}^{N_h}\left|\bar{\sigma}^T(t_j)\widetilde{\mathbf{W}}(t)\right|^{2+\eta} \geq N_h^{-\eta/2}\left(\sum_{j=1}^{N_h}\left|\bar{\sigma}^T(t_j)\widetilde{\mathbf{W}}(t)\right|^2\right)^{1+\eta/2}.\quad (83)$$

By the definition of $\mathbf{\Pi}$ and Assumption 1,

$$\sum_{j=1}^{N_h}\left|\bar{\sigma}^T(t_j)\widetilde{\mathbf{W}}(t)\right|^2 = \widetilde{\mathbf{W}}^T(t)\mathbf{\Pi}\mathbf{\Pi}^T\widetilde{\mathbf{W}}(t) = \widetilde{\mathbf{W}}^T(t)\mathbf{\Sigma}\widetilde{\mathbf{W}}(t) \geq \lambda_0\|\widetilde{\mathbf{W}}(t)\|^2.\quad (84)$$

Moreover, since

$$V_w(t) = \frac{1}{2}\widetilde{\mathbf{W}}^T(t)\mathbf{\Gamma}^{-1}\widetilde{\mathbf{W}}(t) \leq \frac{1}{2\lambda_{\min}(\mathbf{\Gamma})}\|\widetilde{\mathbf{W}}(t)\|^2,\quad (85)$$

one obtains

$$\|\widetilde{\mathbf{W}}(t)\|^2 \geq 2\lambda_{\min}(\mathbf{\Gamma})V_w(t).\quad (86)$$

Substituting (82)–(86) into (81) gives

$$\dot{V}_w(t) \leq -\alpha_1 c_1 V_w^{1-\eta/2}(t) - \alpha_2 c_2 V_w^{1+\eta/2}(t).\quad (87)$$

According to the definitions of α_1 and α_2 in (76),

$$\dot{V}_w(t) \leq -\frac{\pi}{\theta T_w' \eta} \left(V_w(t)^{1-\eta/2} + V_w(t)^{1+\eta/2} \right) \leq -\frac{\pi}{T_w' \eta} \left(V_w(t)^{1-\eta/2} + V_w(t)^{1+\eta/2} \right).\quad (88)$$

By Lemma 1, the ideal critic weight-estimation error converges to zero within the predefined upper bound $T_w = 2T'_w$.

For the nonideal case, let

$$x_0(t) = \bar{\sigma}^T(t) \widetilde{W}(t), \quad x_j(t) = \bar{\sigma}^T(t_j) \widetilde{W}(t), \quad j = 1, \dots, N_h, \quad (89)$$

and

$$r_0(t) = \bar{\varepsilon}_H(t), \quad r_j(t) = \bar{\varepsilon}_H(t_j), \quad j = 1, \dots, N_h. \quad (90)$$

Outside the residual-dominated region, i.e., when $|x_i(t)| \geq |r_i(t)|$, $i = 0, 1, \dots, N_h$, it holds that

$$\text{sgn}(x_i(t)) = \text{sgn}(x_i(t) + r_i(t)), \quad i = 0, 1, \dots, N_h. \quad (91)$$

Therefore, Equation (79) yields

$$\begin{aligned} \dot{V}_w(t) = & -(\alpha_1 + 1) \sum_{i=0}^{N_h} |x_i(t)| |x_i(t) + r_i(t)|^{1-\eta} \\ & - (\alpha_2 + 1) \sum_{i=0}^{N_h} |x_i(t)| |x_i(t) + r_i(t)|^{1+\eta}. \end{aligned} \quad (92)$$

Using the absolute-value inequalities

$$|x_i(t)|^{1-\eta} \leq |x_i(t) + r_i(t)|^{1-\eta} + |r_i(t)|^{1-\eta}, \quad (93)$$

$$|x_i(t)|^{1+\eta} \leq 2^\eta \left(|x_i(t) + r_i(t)|^{1+\eta} + |r_i(t)|^{1+\eta} \right), \quad (94)$$

Equation (92) is upper-bounded by

$$\begin{aligned} \dot{V}_w(t) \leq & -(\alpha_1 + 1) \sum_{i=0}^{N_h} |x_i(t)|^{2-\eta} - 2^{-\eta} (\alpha_2 + 1) \sum_{i=0}^{N_h} |x_i(t)|^{2+\eta} \\ & + (\alpha_1 + 1) \sum_{i=0}^{N_h} |x_i(t)| |r_i(t)|^{1-\eta} + (\alpha_2 + 1) \sum_{i=0}^{N_h} |x_i(t)| |r_i(t)|^{1+\eta}. \end{aligned} \quad (95)$$

Because $\bar{\varepsilon}_H(t)$ and $\bar{\varepsilon}_H(t_j)$ are bounded and the normalized regressors are bounded, there exists a positive constant $\bar{\sigma}_d$ satisfying

$$\max\{\|\bar{\sigma}(t)\|, \|\bar{\sigma}(t_1)\|, \dots, \|\bar{\sigma}(t_{N_h})\|\} \leq \bar{\sigma}_d. \quad (96)$$

Then the residual-induced terms in (95) satisfy

$$\begin{aligned} & (\alpha_1 + 1) \sum_{i=0}^{N_h} |x_i(t)| |r_i(t)|^{1-\eta} + (\alpha_2 + 1) \sum_{i=0}^{N_h} |x_i(t)| |r_i(t)|^{1+\eta} \\ & \leq (N_h + 1) \bar{\sigma}_d \left[(\alpha_1 + 1) \bar{\varepsilon}_{Hd}^{1-\eta} + (\alpha_2 + 1) \bar{\varepsilon}_{Hd}^{1+\eta} \right] \|\widetilde{W}(t)\|. \end{aligned} \quad (97)$$

Let

$$\varrho_w = (N_h + 1) \bar{\sigma}_d \left[(\alpha_1 + 1) \bar{\varepsilon}_{Hd}^{1-\eta} + (\alpha_2 + 1) \bar{\varepsilon}_{Hd}^{1+\eta} \right]. \quad (98)$$

Using the historical data terms and Lemma 2, one obtains

$$\dot{V}_w(t) \leq -\alpha_1 \lambda_0^{1-\eta/2} \|\widetilde{W}(t)\|^{2-\eta} - \alpha_2 N_h^{-\eta/2} \lambda_0^{1+\eta/2} \|\widetilde{W}(t)\|^{2+\eta} + \varrho_w \|\widetilde{W}(t)\|. \quad (99)$$

Define the compact residual set

$$\Omega_w = \left\{ \widetilde{\mathbf{W}} \in \mathbb{R}^M : \|\widetilde{\mathbf{W}}\| \leq \rho_w \right\}, \quad (100)$$

where

$$\rho_w = \max \left\{ \left(\frac{\varrho_w}{(1-\theta)\alpha_1\lambda_0^{1-\eta/2}} \right)^{\frac{1}{1-\eta}}, \left(\frac{\varrho_w}{(1-\theta)\alpha_2N_h^{-\eta/2}\lambda_0^{1+\eta/2}} \right)^{\frac{1}{1+\eta}} \right\}. \quad (101)$$

For $\widetilde{\mathbf{W}}(t) \notin \Omega_w$, inequality (99) implies

$$\dot{V}_w(t) \leq -\theta\alpha_1\lambda_0^{1-\eta/2}\|\widetilde{\mathbf{W}}(t)\|^{2-\eta} - \theta\alpha_2N_h^{-\eta/2}\lambda_0^{1+\eta/2}\|\widetilde{\mathbf{W}}(t)\|^{2+\eta}. \quad (102)$$

Using (86) further yields

$$\dot{V}_w(t) \leq -\theta\alpha_1c_1V_w^{1-\eta/2}(t) - \theta\alpha_2c_2V_w^{1+\eta/2}(t). \quad (103)$$

According to (76), one has

$$\dot{V}_w(t) \leq -\frac{\pi}{T_w'\eta} \left(V_w(t)^{1-\eta/2} + V_w(t)^{1+\eta/2} \right), \quad \widetilde{\mathbf{W}}(t) \notin \Omega_w. \quad (104)$$

Thus, by Lemma 1, $\widetilde{\mathbf{W}}(t)$ converges into the compact set Ω_w within the predefined upper bound $T_w = 2T_w'$ and remains therein thereafter. This completes the proof. $\square$

Theorem 2. For the disturbed physical system (13), apply the RL control policy in (56) and the current-and-history network parameter update law (73). If Assumption 1 holds and the initial constrained errors satisfy (20), then the closed-loop system is uniformly ultimately bounded and the prescribed-performance constraints hold for all $t \geq 0$.

Proof. Define the Lyapunov function

$$V(t) = V_c(\mathbf{z}(t), t) + V_w(\widetilde{\mathbf{W}}(t)). \quad (105)$$

Taking the derivative yields

$$\dot{V}(t) = \dot{V}_c(t) + \dot{V}_w(t). \quad (106)$$

Here,

$$\dot{V}_c(t) = \nabla V_c^T(t)(F(t) + B\hat{\mathbf{u}}(t) + \mathbf{d}(t)) + \frac{\partial V_c}{\partial t} \quad (107)$$

By further manipulation,

$$\begin{aligned} \dot{V}_c(t) &= \frac{\partial V_c(\mathbf{z}(t), t)}{\partial t} + \nabla_z V_c^T(\mathbf{z}(t), t)(F(t) + B\mathbf{u}^*(t) + \mathbf{d}^*(t)) \\ &\quad + \nabla_z V_c^T(\mathbf{z}(t), t)B(\hat{\mathbf{u}}(t) - \mathbf{u}^*(t)) + \nabla_z V_c^T(\mathbf{z}(t), t)(\mathbf{d}(t) - \mathbf{d}^*(t)) \end{aligned} \quad (108)$$

According to the optimality condition,

$$\begin{aligned} &\frac{\partial V_c(\mathbf{z}(t), t)}{\partial t} + \nabla_z V_c^T(\mathbf{z}(t), t)(F(t) + B\mathbf{u}^*(t) + \mathbf{d}^*(t)) \\ &= -\mathbf{z}^T(t)Q\mathbf{z}(t) - \mathbf{u}^{*T}(t)R\mathbf{u}^*(t) - \mathcal{B}_p(\mathbf{z}(t), t) + \delta^2\mathbf{d}^{*T}(t)\mathbf{d}^*(t). \end{aligned} \quad (109)$$

Moreover, from the extremum conditions of the optimal control and disturbance laws,

$$\nabla V_c^T(t)B = -2\mathbf{u}^{*T}(t)R, \quad \nabla V_c^T(t) = 2\delta^2\mathbf{d}^{*T}(t). \quad (110)$$

one has

$$\begin{aligned}\dot{V}_c(t) = & -\mathbf{z}^T(t)\mathbf{Q}\mathbf{z}(t) - \mathbf{u}^{*\top}(t)\mathbf{R}\mathbf{u}^*(t) - \mathcal{B}_p(\mathbf{z}(t), t) + \delta^2\mathbf{d}^{*\top}(t)\mathbf{d}^*(t) \\ & - 2\mathbf{u}^{*\top}(t)\mathbf{R}(\hat{\mathbf{u}}(t) - \mathbf{u}^*(t)) + 2\delta^2\mathbf{d}^{*\top}(t)(\mathbf{d}(t) - \mathbf{d}^*(t))\end{aligned}\quad (111)$$

Noting that

$$-\mathbf{u}^{*\top}(t)\mathbf{R}\mathbf{u}^*(t) - 2\mathbf{u}^{*\top}(t)\mathbf{R}(\hat{\mathbf{u}}(t) - \mathbf{u}^*(t)) = -\hat{\mathbf{u}}^T(t)\mathbf{R}\hat{\mathbf{u}}(t) + (\hat{\mathbf{u}}(t) - \mathbf{u}^*(t))^T\mathbf{R}(\hat{\mathbf{u}}(t) - \mathbf{u}^*(t)) \quad (112)$$

$$\delta^2\mathbf{d}^{*\top}(t)\mathbf{d}^*(t) + 2\delta^2\mathbf{d}^{*\top}(t)(\mathbf{d}(t) - \mathbf{d}^*(t)) = \delta^2\mathbf{d}^T(t)\mathbf{d}(t) + \delta^2(\mathbf{d}(t) - \mathbf{d}^*(t))^T(\mathbf{d}(t) - \mathbf{d}^*(t)) \quad (113)$$

then, one has

$$\begin{aligned}\dot{V}_c(t) \leq & -\mathbf{z}^T(t)\mathbf{Q}\mathbf{z}(t) - \mathcal{B}_p(\mathbf{z}(t), t) - \hat{\mathbf{u}}^T(t)\mathbf{R}\hat{\mathbf{u}}(t) + \delta^2\mathbf{d}^T(t)\mathbf{d}(t) \\ & + (\hat{\mathbf{u}}(t) - \mathbf{u}^*(t))^T\mathbf{R}(\hat{\mathbf{u}}(t) - \mathbf{u}^*(t)) \\ & - \delta^2(\mathbf{d}(t) - \mathbf{d}^*(t))^T(\mathbf{d}(t) - \mathbf{d}^*(t))\end{aligned}\quad (114)$$

Combining the critic neural-network approximation property, the errors between the practical and optimal policies are expressed by the parameter reconstruction error and the approximation residual as

$$\hat{\mathbf{u}}(t) - \mathbf{u}^*(t) = -\frac{1}{2}\mathbf{R}^{-1}\mathbf{B}^T\left(\nabla\psi^T(\mathbf{z}(t))\widetilde{\mathbf{W}}(t) - \nabla\varepsilon(\mathbf{z}(t))\right), \quad (115)$$

Theorem 1 has shown that $\widetilde{\mathbf{W}}(t)$ converges into a bounded region within the predefined time. On a compact set, $\|\widetilde{\mathbf{W}}(t)\| \leq \rho_w$, $\|\nabla\psi(\mathbf{z}(t))\| \leq \bar{\psi}_d$, $\|\nabla\varepsilon(\mathbf{z}(t))\| \leq \varepsilon_{dc}$, and $\|\mathbf{d}(t)\| \leq \bar{d}$, then

$$\|\hat{\mathbf{u}} - \mathbf{u}^*\| \leq \frac{1}{2}\|\mathbf{R}^{-1}\mathbf{B}^T\|(\bar{\psi}_d\rho_w + \varepsilon_{dc}) = \bar{u}_e$$

Therefore, define a positive constant D_M such that the residual terms induced by the network approximation error, and the external bounded disturbance satisfy

$$\delta^2\mathbf{d}^T(t)\mathbf{d}(t) + (\hat{\mathbf{u}}(t) - \mathbf{u}^*(t))^T\mathbf{R}(\hat{\mathbf{u}}(t) - \mathbf{u}^*(t)) \leq D_M \quad (116)$$

Using the property of matrix quadratic forms $\mathbf{x}^T\mathbf{M}\mathbf{x} \leq \lambda_{\max}(\mathbf{M})\|\mathbf{x}\|^2$, we explicitly construct the constant D_M as

$$D_M = \lambda_{\max}(\mathbf{R})\bar{u}_e^2 + \delta^2\bar{d}^2 \quad (117)$$

The quadratic form $-\mathbf{z}^T(t)\mathbf{Q}\mathbf{z}(t)$ satisfies

$$\dot{V}_c(t) \leq -\lambda_{\min}(\mathbf{Q})\|\mathbf{z}(t)\|^2 + D_M. \quad (118)$$

Substituting this result into the derivative of the composite Lyapunov functional and using Theorem 1 gives, for $\widetilde{\mathbf{W}}(t) \notin \Omega_w$,

$$\dot{V}(t) \leq -\lambda_{\min}(\mathbf{Q})\|\mathbf{z}(t)\|^2 + D_M - \frac{\pi}{T'_w\eta}\left(V_w(t)^{1-\eta/2} + V_w(t)^{1+\eta/2}\right). \quad (119)$$

When $\widetilde{\mathbf{W}}(t) \in \Omega_w$, the critic error is already bounded. Therefore, when

$$\|\mathbf{z}(t)\| > \sqrt{\frac{D_M}{\lambda_{\min}(\mathbf{Q})}}, \quad (120)$$

one has $\dot{V}(t) < 0$. According to Lyapunov stability theory, the coupled closed-loop system is uniformly ultimately bounded.

It remains to establish satisfaction of the original prescribed-performance constraints. Let $[0, t_{\max})$ be the maximal interval on which all direct and recursive margins are positive. From (51) and the boundedness established above, there exists a constant $\bar{\mathcal{B}}_p > 0$ such that

$$0 \leq \mathcal{B}_p(z(t), t) \leq |V_c^*(z(t), t)| + \|\mathbf{W}^*\| \|\boldsymbol{\psi}(z(t))\| + |\varepsilon(z(t), t)| \leq \bar{\mathcal{B}}_p, \quad 0 \leq t < t_{\max}. \quad (121)$$

For a directly actuated channel $i \in \mathcal{D}_p$, one has $0 \leq \mathcal{B}_i^d \leq \mathcal{B}_p \leq \bar{\mathcal{B}}_p$. Suppose that $\phi_{i,U}$ first reaches zero at a finite $t_{\max}$. Since $\phi_{i,L}(t_{\max}) = 2\sigma_i(t_{\max}) > 0$, it follows that

$$\mathcal{B}_i^d(z(t), t) = k_{d,i} \left[\ln \left(\frac{\phi_{i,U}(t)}{\phi_{i,L}(t)} \right) \right]^2 \rightarrow +\infty, \quad t \rightarrow t_{\max}^-, \quad (122)$$

which contradicts the boundedness of $\mathcal{B}_p$. The same argument applies if $\phi_{i,L}$ reaches zero. Therefore,

$$\phi_{i,U}(t) > 0, \quad \phi_{i,L}(t) > 0, \quad i \in \mathcal{D}_p. \quad (123)$$

For a higher-relative-degree channel $i \in \mathcal{H}_p$, define $a_i(t) = \dot{\sigma}_i(t) + \lambda_i \sigma_i(t) - \bar{\Delta}_i$ and suppose that $a_i(t) \geq \underline{a}_i > 0$. The recursive-barrier component satisfies

$$0 \leq \mathcal{B}_{i,1}^h(z, t) = k_{1,i} \left[\ln \left(\frac{r_{i,U}(t)}{r_{i,L}(t)} \right) \right]^2 \leq \mathcal{B}_p(z, t) \leq \bar{\mathcal{B}}_p. \quad (124)$$

Since $r_{i,U}(t) + r_{i,L}(t) = 2a_i(t) \geq 2\underline{a}_i > 0$, if $r_{i,U}$ first reaches zero, then $r_{i,L} \geq 2\underline{a}_i > 0$ and $\mathcal{B}_{i,1}^h \rightarrow +\infty$, which is impossible. The same argument applies to $r_{i,L}$. Hence,

$$r_{i,U}(t) > 0, \quad r_{i,L}(t) > 0, \quad i \in \mathcal{H}_p. \quad (125)$$

By the robust recursive-boundary construction, $r_{i,U} > 0$ and $r_{i,L} > 0$ imply

$$\dot{\phi}_{i,U}(t) + \lambda_i \phi_{i,U}(t) > 0, \quad \dot{\phi}_{i,L}(t) + \lambda_i \phi_{i,L}(t) > 0. \quad (126)$$

Using the initial conditions $\phi_{i,U}(0) > 0$ and $\phi_{i,L}(0) > 0$ gives

$$\phi_{i,U}(t) > e^{-\lambda_i t} \phi_{i,U}(0) > 0, \quad \phi_{i,L}(t) > e^{-\lambda_i t} \phi_{i,L}(0) > 0, \quad i \in \mathcal{H}_p. \quad (127)$$

Consequently, $t_{\max} = \infty$ and

$$-\sigma_i(t) < z_i(t) < \sigma_i(t), \quad i \in \mathcal{D}_p \cup \mathcal{H}_p, \quad t \geq 0. \quad (128)$$

Thus, boundedness of the direct barrier guarantees the constraints for the directly actuated channels, whereas boundedness of the recursive barrier guarantees the constraints for the higher-relative-degree channels through the recursive boundary dynamics. This completes the proof. $\square$

5. Simulation

To thoroughly assess the effectiveness of ZG-RL-PP, this section presents a systematic experimental study aimed at evaluating its generalizability, performance, and robustness. Specifically, the validation framework is structured into two components.

5.1. Comparison with Traditional Methods in Different Trim Points

In this section, three controllers are compared in all response figures. ZG-RL-PP denotes the proposed controller with the prescribed-performance barrier gradient. ZG-RL

uses the same zero-sum critic update and the same saturation/disturbance settings, but removes the barrier gradient from the policy. IRL denotes a fault-free nominal integral-reinforcement-learning benchmark adapted from the baseline method in [34]; the actuator-fault model and the fault-compensation terms are removed, and the system matrices, reference command, initial state, disturbance, and input saturation are kept identical to the present simulation. The purpose of the IRL curve is therefore to provide a nominal data-driven tracking baseline rather than a fault-tolerant comparison.

The work range of the linear model is when the dihedral angle is 0 degree to 15 degrees [33]. To validate the generalization capability of the proposed method across the model, two trim points with distinct dihedral angles of 15 degrees and 5 degrees were selected, representing relatively large and small configurations. The HFA is trimmed under the following conditions: $V = 68 \text{ ft/s}$, $h = 40,000 \text{ ft}$, $\alpha = 2.8^\circ$, $\theta = 2.8^\circ$, $\eta = 15^\circ$ or 5° , $\dot{\eta} = 0$.

The two HFA trim points are denoted as System 1 and System 2. System 1 uses (A_1, B_1) , where

$$A_1 = \begin{bmatrix} -0.0428 & 4.1764 & -32.2 & -0.0010 & 0.2557 & -0.0292 \\ -0.0139 & -8.7716 & 0 & 1.0045 & 0.1618 & 0.1564 \\ 0 & 0 & 0 & 1.0000 & 0 & 0 \\ 0 & -71.3347 & 0 & -0.0052 & 0.8590 & 8.2122 \\ 0 & 0 & 0 & 0 & 0 & 1.0000 \\ 0.0015 & 0.3995 & 0 & 1.9228 & -0.0308 & -7.5587 \end{bmatrix}, \quad (129)$$

$$B_1 = \begin{bmatrix} 0.0357 & -2.6729 & -6.4431 & 0 & 0 \\ 0 & -0.9144 & -1.7666 & -0.1795 & -0.3469 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -2.2479 & -3.9797 & -16.3906 & -33.4039 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -0.8925 & 1.0479 & -0.1749 & 0.2053 \end{bmatrix}. \quad (130)$$

System 2 uses (A_2, B_2) , where

$$A_2 = \begin{bmatrix} -0.0417 & 4.1722 & -32.2 & -0.0008 & 0.0835 & 0.1788 \\ -0.0139 & -9.1315 & 0 & 1.0009 & 0.0541 & 0.0191 \\ 0 & 0 & 0 & 1.0000 & 0 & 0 \\ 0 & -446.4040 & 0 & -0.1753 & 2.2393 & 38.4633 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0.0005 & 0.0480 & 0 & 1.9991 & -0.1384 & -7.4605 \end{bmatrix}, \quad (131)$$

$$B_2 = \begin{bmatrix} 0.0357 & -2.6729 & -6.2501 & 0 & 0 \\ 0 & -0.9144 & -1.8219 & -0.1795 & -0.3577 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -12.8245 & -24.5028 & -91.5144 & -192.2739 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -0.9196 & 0.9375 & -0.1802 & 0.1836 \end{bmatrix}. \quad (132)$$

The bounded disturbance used in the plant propagation is

$$\mathbf{d}(t) = 1.1[0.1 \sin(0.1t), 0.2 \sin(0.1t), 0.2 \cos(0.1t), 0.1 \sin(0.1t), 0.1 \cos(0.1t), 0.1 \sin(0.1t)]^T. \quad (133)$$

The disturbance used in the reported simulations is $\mathbf{d}(t)$, which is independently defined; it is neither the exact worst-case disturbance $\mathbf{d}^*(t)$ nor its estimate $\hat{\mathbf{d}}(t)$. The zero-sum game weights and critic-learning parameters are selected as

$$Q = 10 \operatorname{diag}(100, 100, 0.001, 0.001, 100, 0.001), \quad R = 80 \operatorname{diag}(1, 280, 240, 5, 5), \quad \delta = 200. \quad (134)$$

The learning time is $T_w = 10$ s, the fractional parameter is $\eta = 0.8$, the number of basis functions is $M = 25$, the critic gain matrix is $\Gamma = 0.01 I_M$, and the initial critic weight is $\hat{W}(0) = 500 I_M$. The simulation step is 10^{-4} s, the final time is 20 s, and the initial state is $x(0) = [5, -0.05, 0, 0, 0, 0]^T$. The reference command and the auxiliary command-rate signal used in the cross-basis functions are

$$x_d(t) = [7 + 0.01t + 2 \sin(0.5t), 0.1, 0, 0, 0.5, 0]^T, \quad (135)$$

$$\dot{x}_d(t) = [0.01 + \cos(0.5t), 0, 0, 0, 0, 0]^T. \quad (136)$$

For the prescribed-performance controller, the directly constrained channels are z_1 and z_2 , and the elastic coordinate z_5 is treated as a high-relative-degree constrained channel through z_6 . The direct-channel prescribed time is $T_c = 2.5$ s, and the high-relative-degree prescribed time is $T_{c,hr} = 5.5$ s. The envelope parameters are $\alpha_\sigma = 0.5$, $\sigma_0 = [3, 0.5, 0.5, 0.5, 2, 0.5]^T$, and $\sigma_\infty = 0.0021_6$. The direct barrier gains are selected as $K_{\text{perf}} = 0.4[10, 50, 0, 0, 5, 0]^T$. For the high-relative-degree channel, the recursive-margin parameters are $\lambda_5 = 5000$, $k_{51} = 50000$, $\bar{\Delta}_5 = 0.65$, and $\varepsilon_r = 10^{-8}$.

Figures 3 and 4 show the state responses for the two systems. In both systems, the ZG-RL-PP controller tracks the velocity command and drives the angle-of-attack and elastic-coordinate responses to the desired values after the initial transient. The ZG-RL controller reaches comparable state tracking for the main channels, but it does not impose the transient error envelope. The nominal IRL benchmark follows the velocity command after a short transient, but its angle-of-attack and elastic-coordinate responses exhibit slower error removal and larger steady deviation, especially in the elastic coordinate.

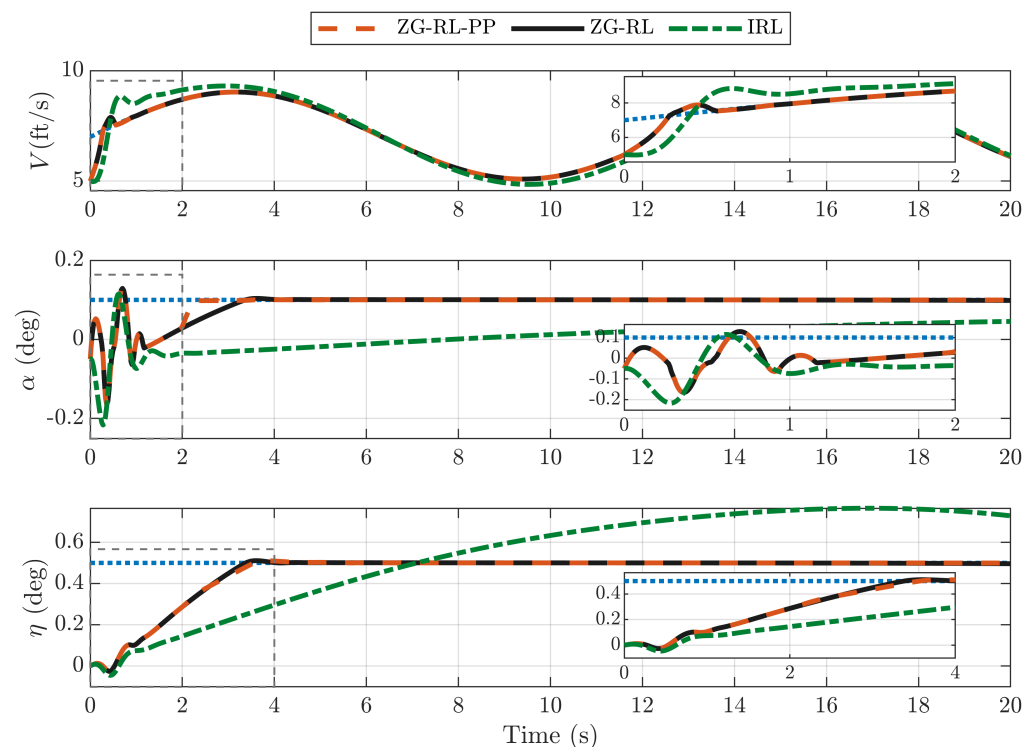


Figure 3. State tracking comparison for System 1.

Figures 5 and 6 report the tracking errors together with the prescribed-performance envelopes and local magnifications around the prescribed times and the terminal interval.

The With PP errors remain inside the shrinking envelopes for z_1 , z_2 , and z_5 . In contrast, the without PP and nominal IRL responses are not constrained by the barrier function, and their local magnifications show visible envelope violations or larger terminal errors. These comparisons directly illustrate the role of the proposed barrier terms: the barrier does not merely improve the final tracking error, but also enforces the prescribed transient and terminal admissible sets.

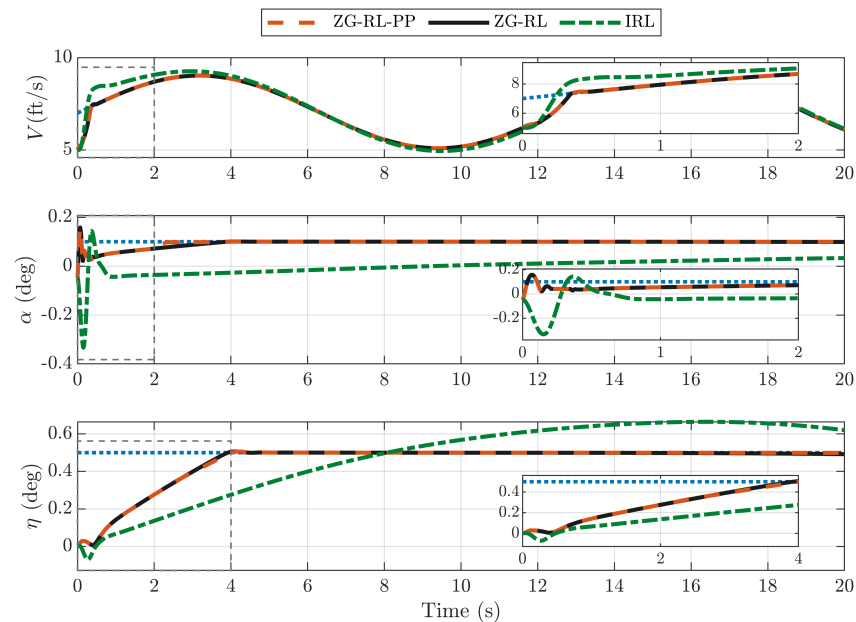


Figure 4. State tracking comparison for System 2.

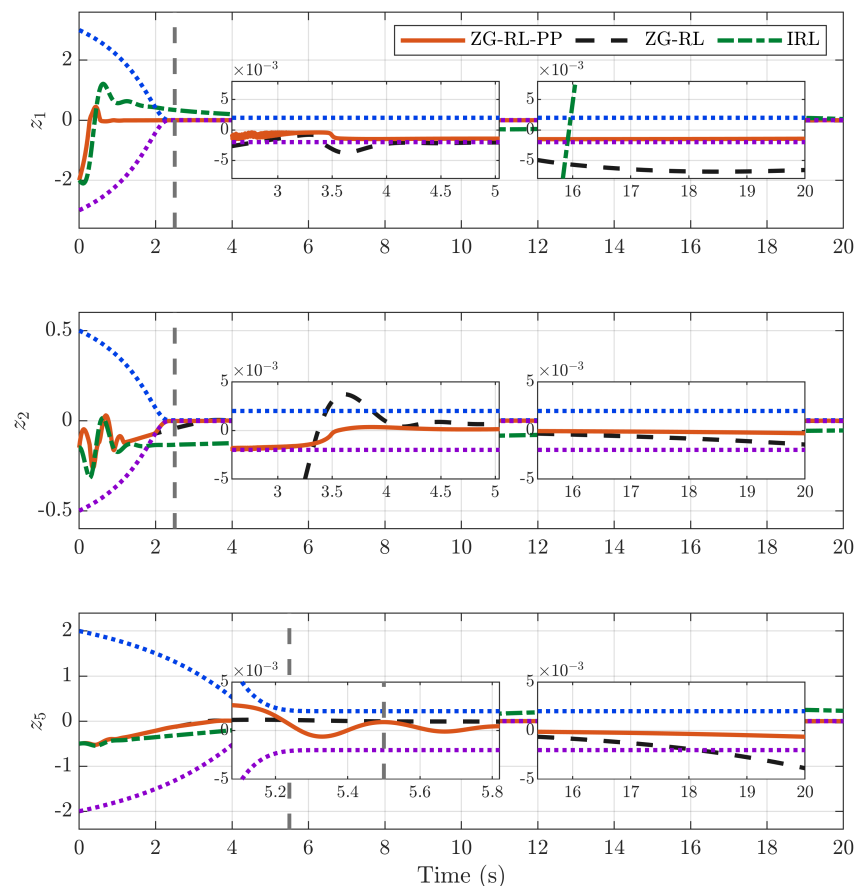


Figure 5. Tracking-error and prescribed-performance comparison for System 1.

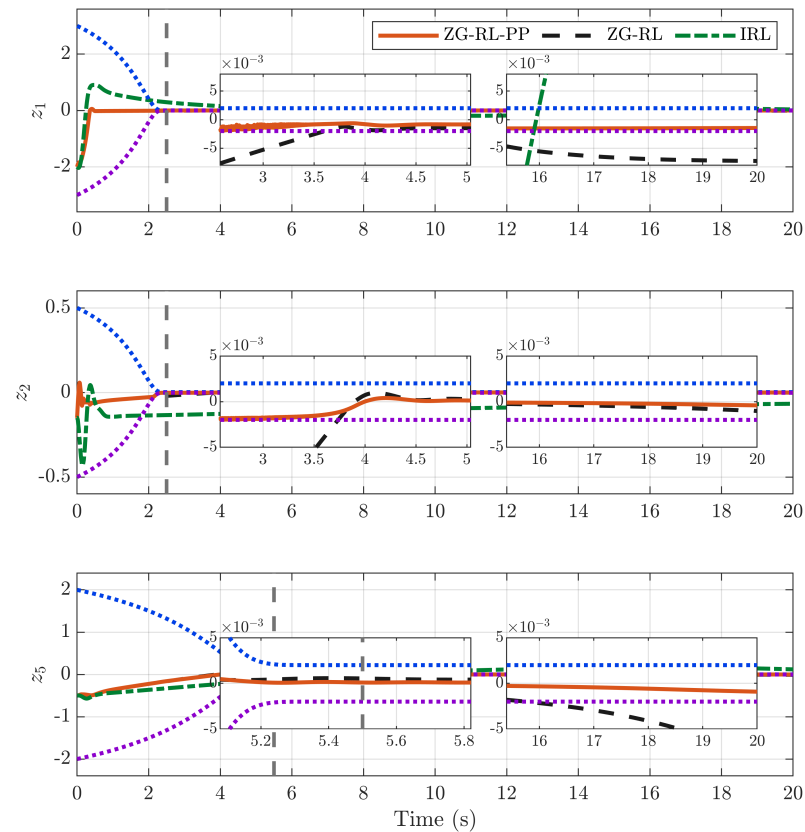


Figure 6. Tracking-error and prescribed-performance comparison for System 2.

Figures 7 and 8 compare the five input channels. The ZG-RL-PP controller uses additional control activity during the transient interval to maintain the prescribed-performance constraints, while the nominal IRL benchmark generates smaller nominal commands because it does not contain the zero-sum disturbance-attenuation term or the performance barrier. All input channels remain bounded under the imposed saturation structure.

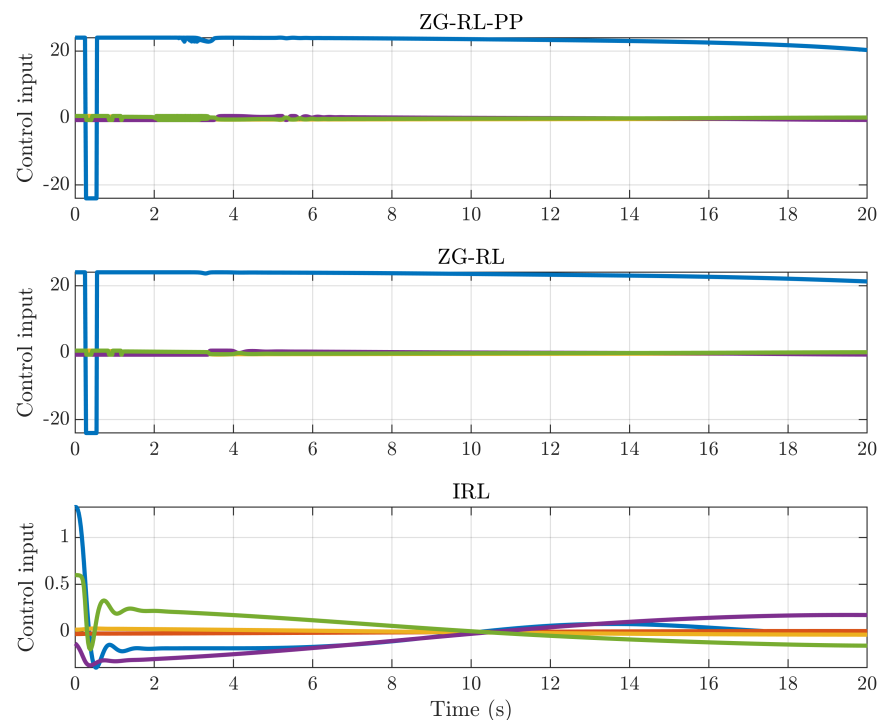


Figure 7. Control-input comparison for System 1.

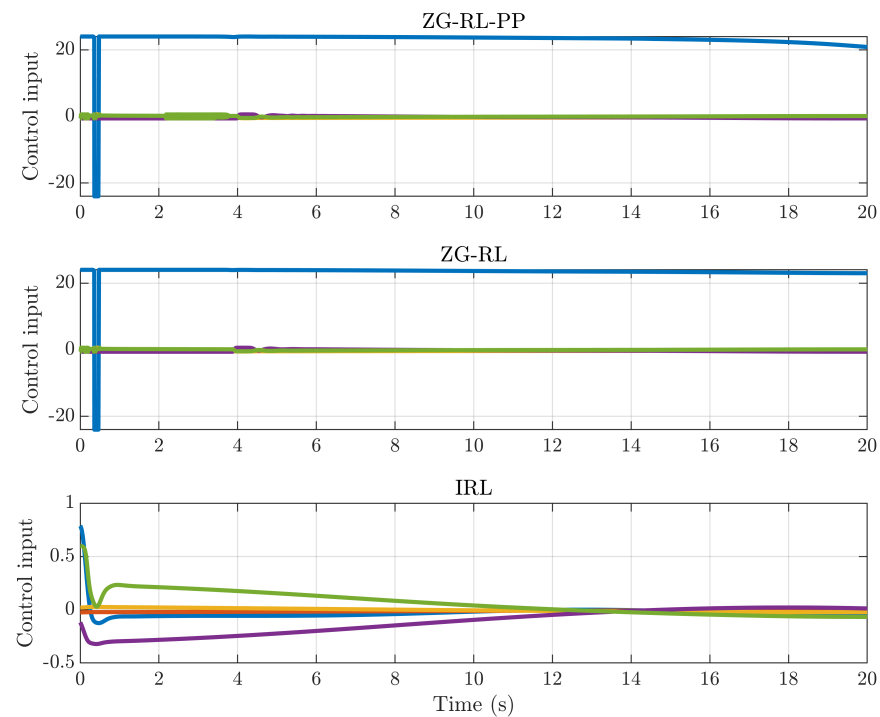


Figure 8. Control-input comparison for System 2.

Figures 9 and 10 show the Bellman residual and the weight-update norm on logarithmic scales for the two critic-based zero-sum controllers. Both quantities decrease after the dominant learning transient and remain bounded. The prescribed-performance terms change the transient learning trajectory because the barrier gradient enters the policy and the Bellman residual, but learning activity is still concentrated before and around the assigned learning window.

Figures 11 and 12 show the critic neural-network weights and the norm of the weight increment. The weights experience a fast adjustment during the early learning interval and remain bounded afterward. The bounded evolution of $\|\Delta W\|_2$ is consistent with the practical predefined-time critic convergence analysis. The IRL benchmark is not included in these two figures because it is implemented as a fixed nominal gain benchmark rather than as the same online critic neural network.

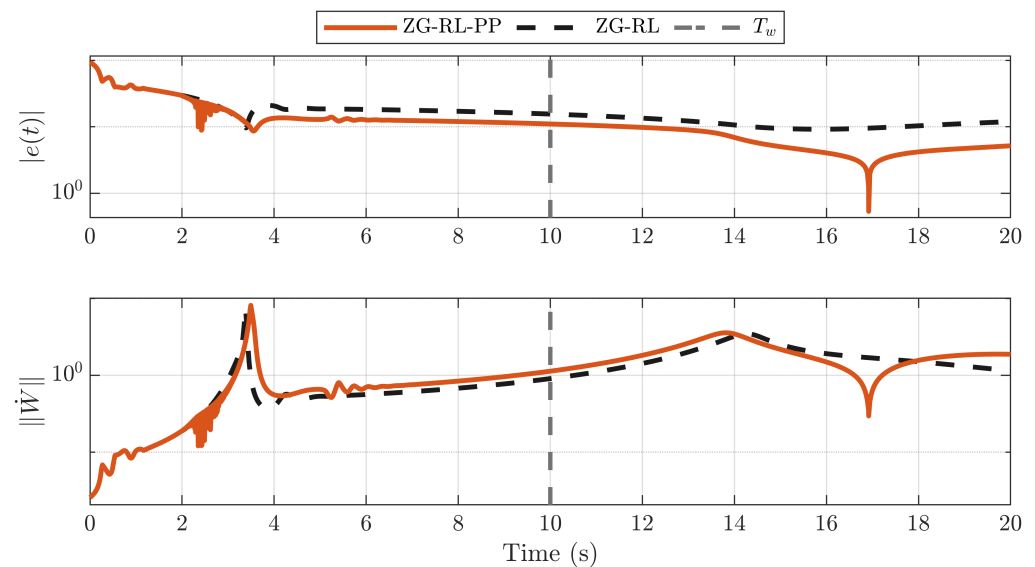


Figure 9. Critic learning indicators for System 1.

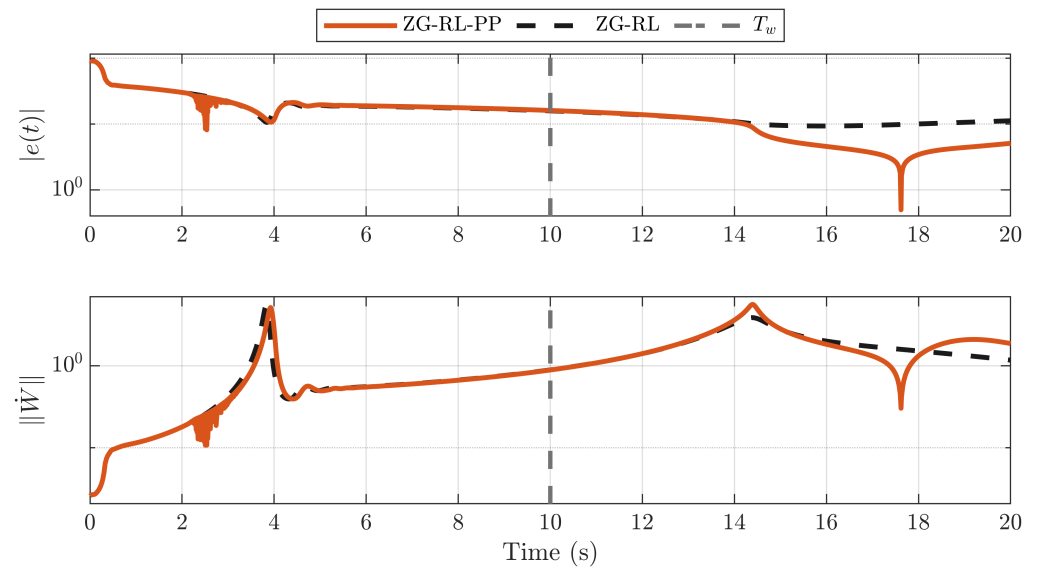


Figure 10. Critic learning indicators for System 2.

Tables 1 and 2 show that only ZG-RL-PP reaches the terminal performance bounds for all three tracking channels without constraint violation. For System 1, ZG-RL-PP achieves the lowest RMSE and maximum absolute error in all channels. For System 2, it achieves the lowest RMSE for z_1 and z_2 and comparable performance for z_5 , while its maximum errors are no larger than those of ZG-RL and are lower than those of IRL. This improved constraint satisfaction and convergence performance is obtained with control energy comparable to ZG-RL.

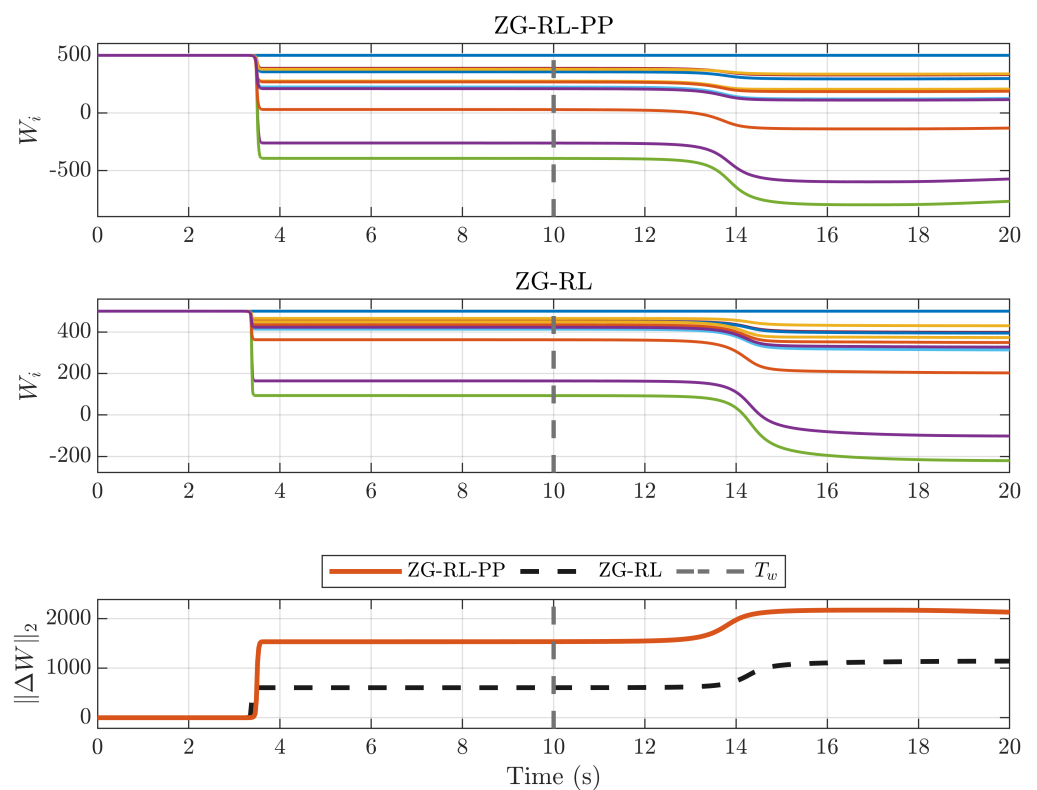


Figure 11. Critic neural-network weights for System 1.

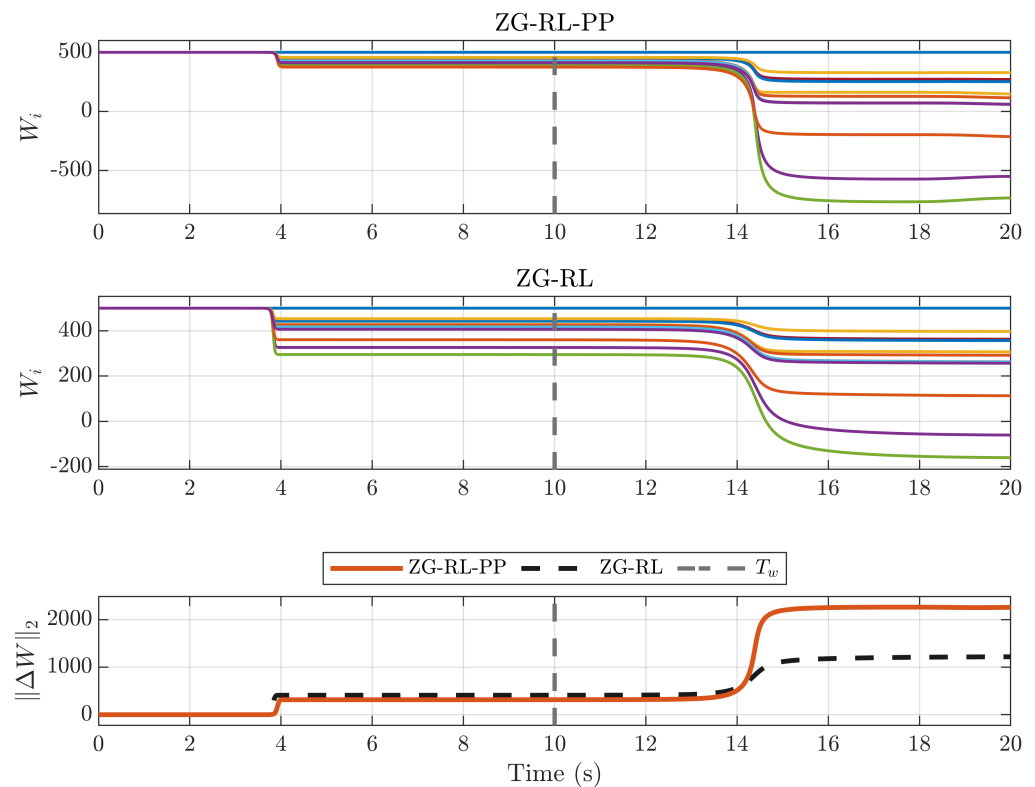


Figure 12. Critic neural-network weights for System 2.

Table 1. Quantitative performance comparison for system 1.

Method	RMSE			Max. Absolute Error			Convergence Time (s)			Control Energy	Constraint Violation
	z_1	z_2	z_5	z_1	z_2	z_5	z_1	z_2	z_5		
ZG-RL	0.05230	0.07471	0.06471	0.66667	0.53029	0.26340	N.R.	3.8698	N.R.	9.5093×10^5	Yes
ZG-RL-PP	0.01655	0.06798	0.05923	0.33333	0.40250	0.25000	2.3560	2.3673	5.1710	9.4322×10^5	No
IRL	0.11873	0.20134	0.12036	0.70264	0.63321	0.27309	N.R.	N.R.	N.R.	723.28	Yes

N.R. indicates that the corresponding terminal performance bound was not reached within the simulation interval.

Table 2. Quantitative performance comparison for System 2.

Method	RMSE			Max. Absolute Error			Convergence Time (s)			Control Energy	Constraint Violation
	z_1	z_2	z_5	z_1	z_2	z_5	z_1	z_2	z_5		
ZG-RL	0.06291	0.03153	0.06285	0.66667	0.30000	0.25000	N.R.	3.7363	N.R.	9.8375×10^5	Yes
ZG-RL-PP	0.06290	0.03044	0.06288	0.66667	0.30000	0.25000	2.4442	2.4410	4.4312	9.7371×10^5	No
IRL	0.09294	0.21460	0.10257	0.67992	0.86657	0.28544	N.R.	N.R.	N.R.	527.35	Yes

N.R. indicates that the corresponding terminal performance bound was not reached within the simulation interval.

Three groups of initial conditions are adopted for system (A_1, B_1) : Case 1: $x(0) = [8, -0.05, 0, 0, 0, 0]^T$, Case 2: $x(0) = [4.6, -0.06, 0, 0, 0, 0]^T$, and Case 3: $x(0) = [4.2, -0.04, 0, 0, 0, 0]^T$. Figure 13 collects the comparative simulation results. Subfigures (a), (c), (e) illustrate the tracking-error responses corresponding to three initial states, while subfigures (b), (d), (f) depict the time-varying trajectories of weight W .

From Figure 13, it can be observed that, for all three initial-condition cases, the tracking errors converge to their prescribed terminal neighborhoods within the corresponding predefined times. Moreover, consistent with Theorem 1, the critic-weight trajectories converge to bounded steady-state neighborhoods within the predefined learning time T_w , demonstrating practical predefined-time convergence under the tested initial conditions.

Overall, the simulation results verify four points. First, the proposed zero-sum controller preserves bounded tracking under the bounded disturbance for both HFA parameter

sets. Second, adding the prescribed-performance barrier enforces the predefined transient and terminal error envelopes, whereas ZG-RL and nominal IRL baselines do not guarantee this behavior. Third, the online critic weights and Bellman residuals remain bounded under the prescribed-time update law. Fourth, the same controller structure works for both systems, indicating robustness to the considered model-parameter variation.

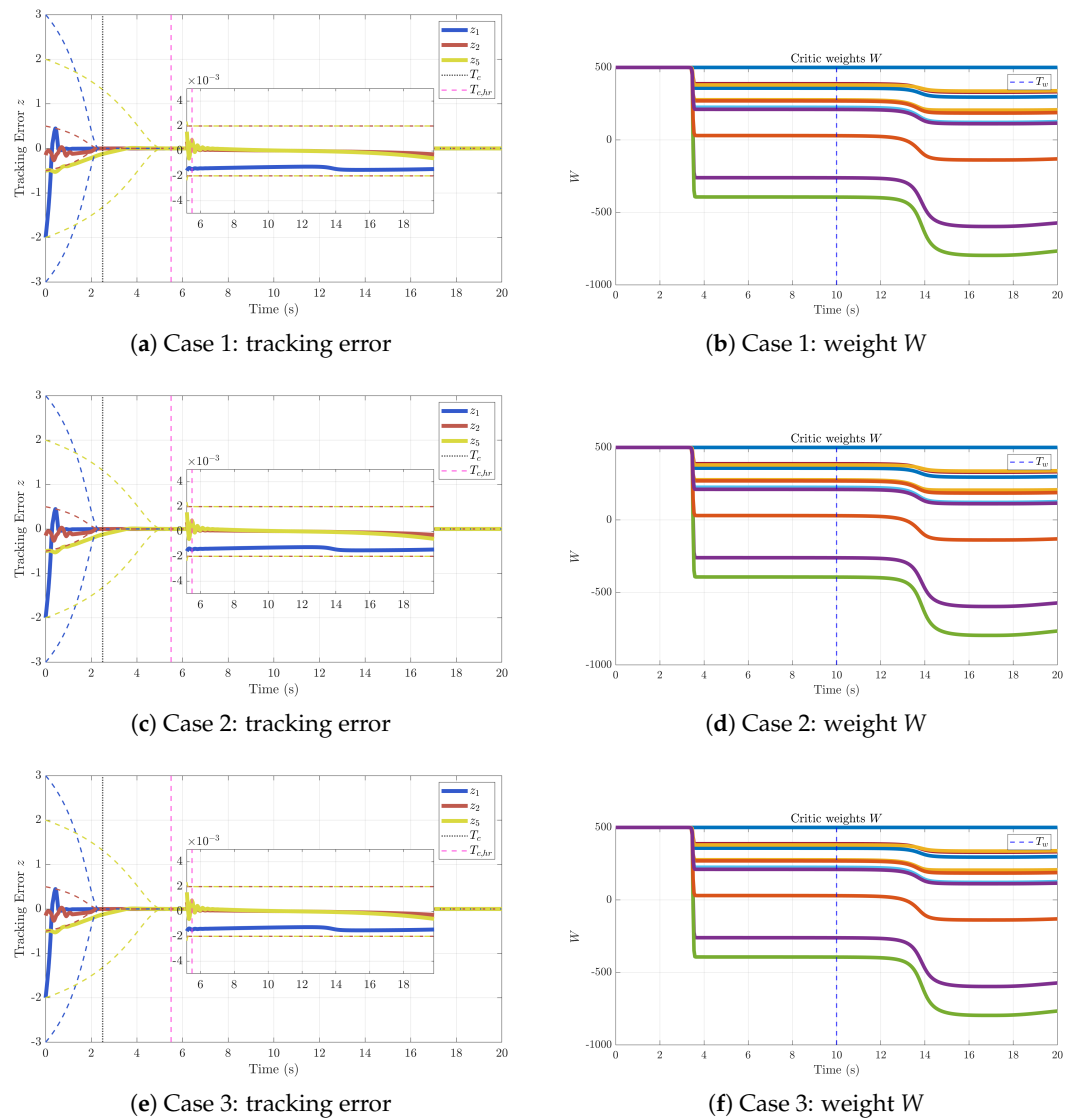


Figure 13. Simulation results of tracking error and weight W under three different initial states.

5.2. Robustness Evaluation Under Model Uncertainty and Measurement Noise

In this section, the robustness of ZG-RL-PP will be verified by model uncertainty simulation. The linearized state matrix and input matrix is (A_2, B_2) appeared in the previous section. The linearized system has eigenvalues with $\lambda_1 = -7.11$, $\lambda_2 = -0.019$, $\lambda_{3,4} = -0.016 \pm 0.64i$, $\lambda_{5,6} = -4.82 \pm 18.66i$ [34]. The system matched uncertainties are caused by flexible effects of dihedral angle [35].

In the simulation environment, A_{real} replaces A in the calculation of environmental information feedback, when the proposed algorithm has no information about it and still uses A in the calculation of the controller, where

$$A_{real} = A + \Delta A \quad (137)$$

Different ΔA are settled as follows:

Case I: no model uncertainty, $\Delta A = 0$.

Case II: low model uncertainty, $\Delta A = 0.2A$.

Case III: high model uncertainty, $\Delta A = 0.5A$.

Case IV: negative model uncertainty, $\Delta A = -0.3A$.

Other simulation conditions are the same as the previous section.

Figures 14–16 further assess the robustness of ZG-RL-PP under model uncertainties. As shown in Figure 14, the tracking errors remain bounded and converge to a small neighborhood of zero in all four cases. Moreover, Figure 17 shows that comparable transient responses and steady-state tracking accuracies are achieved as ΔA varies from $0.5A$ to $-0.3A$, indicating that satisfactory tracking performance is maintained in the presence of model uncertainty.

The sensor measurement noisy is given by

$$x_m(t) = x(t) + n(t), \quad (138)$$

where $n(t) = [0.02 \cos(0.1t), 0.005 \sin(0.5t), 0.005 \sin(0.1t), 0.005 \cos(0.5t), 0.005 \sin(0.1t), 0.005 \cos(0.5t)]^T$ is the measurement noise. The obtained tracking errors and critic-weight responses are provided below.

Figures 18 and 19 show simulation outcomes under measurement noise. Tracking errors stay bounded and converge to a small zero-neighborhood after short transients. The critic weights exhibit bounded learning dynamics and converge to steady values before T_w . Hence, tracking performance is not severely impaired for all tested measurement-noise amplitudes.

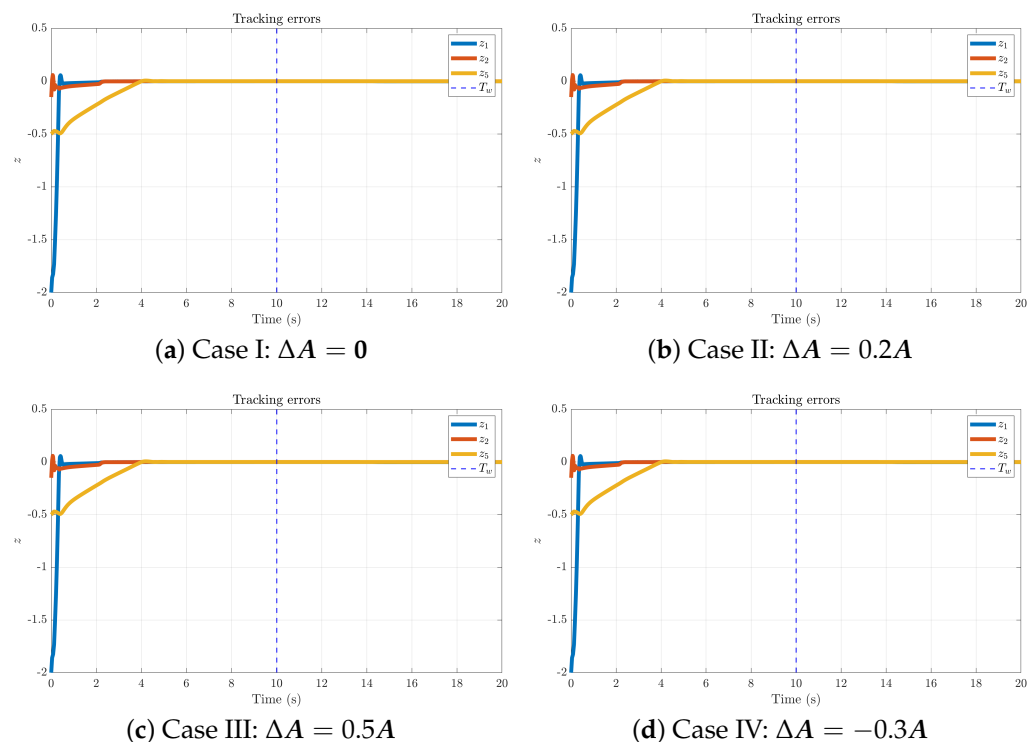


Figure 14. Comparison of model uncertainties regarding error tracking. (a) Case I. (b) Case II. (c) Case III. (d) Case IV.

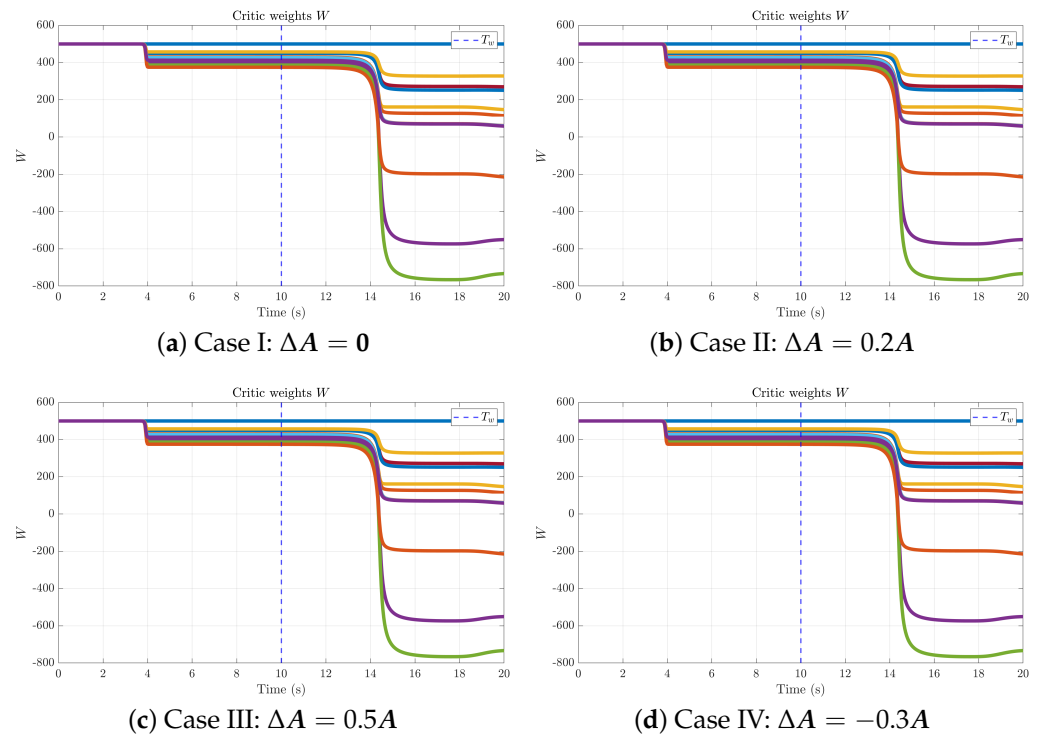


Figure 15. Comparison of model uncertainties regarding critic weights. (a) Case I. (b) Case II. (c) Case III. (d) Case IV.

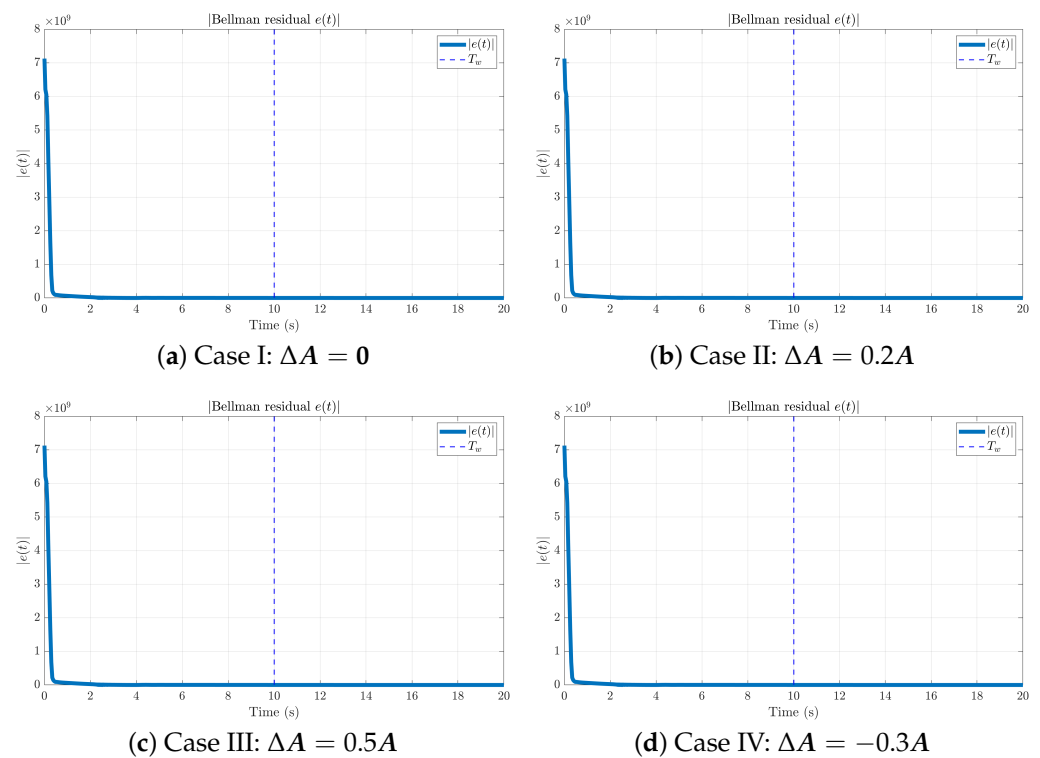


Figure 16. Comparison of model uncertainties regarding Bellman residuals. (a) Case I. (b) Case II. (c) Case III. (d) Case IV.

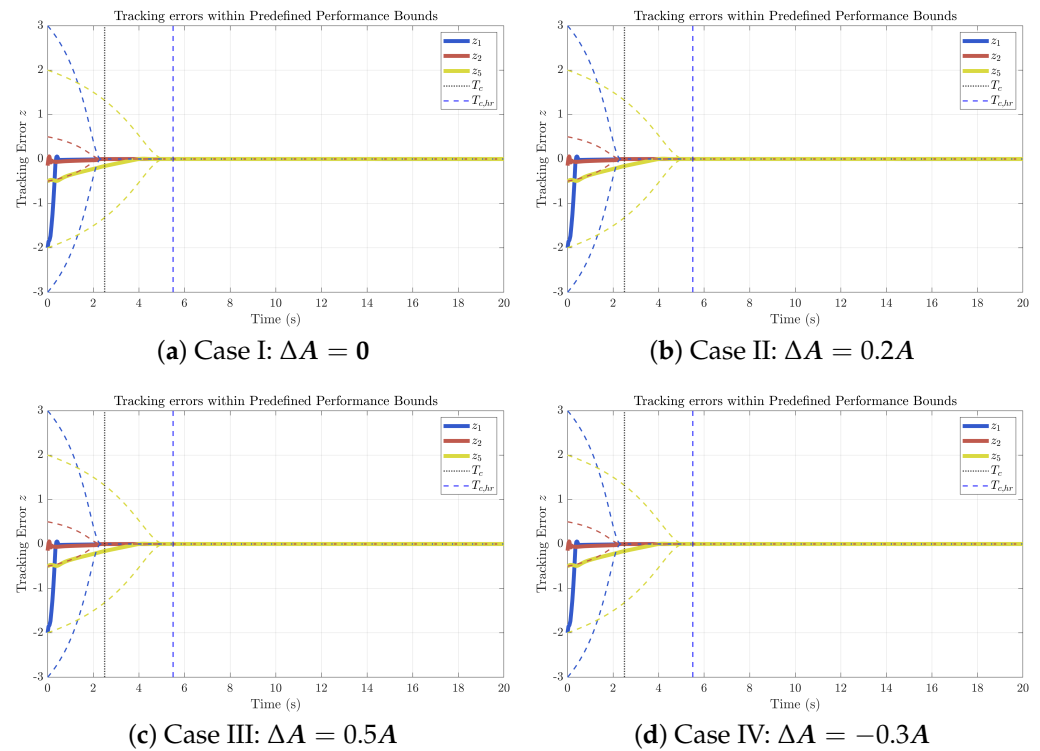


Figure 17. Comparison of model uncertainties regarding tracking errors with bounds. (a) Case I. (b) Case II. (c) Case III. (d) Case IV.

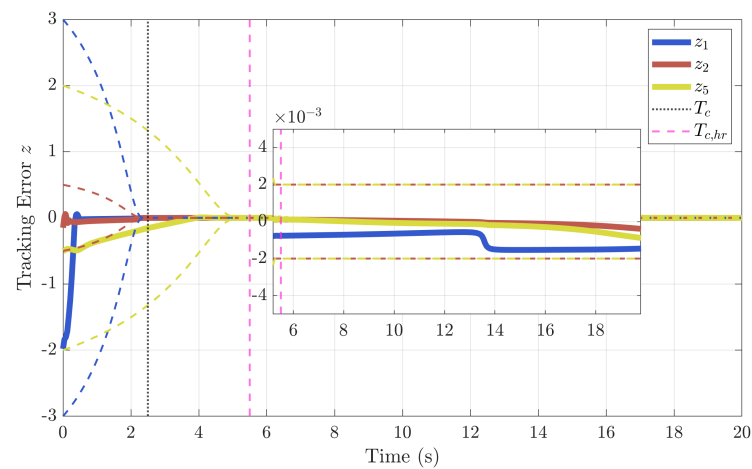


Figure 18. Time response of tracking errors.

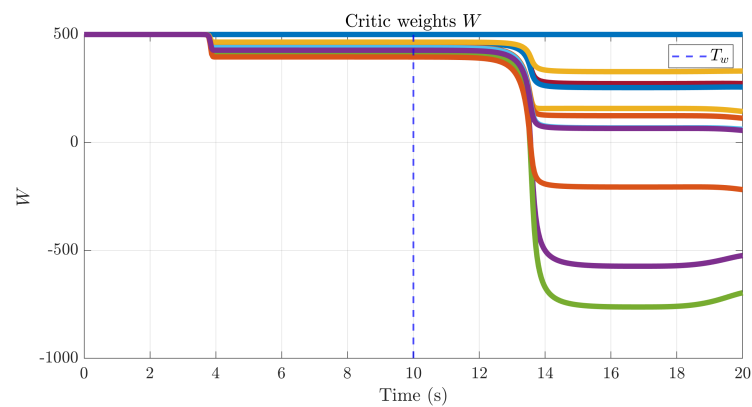


Figure 19. Time response of critic weights.

6. Conclusions

This paper has investigated zero-sum game-based reinforcement learning tracking control for highly flexible aircraft under predefined-time prescribed-performance constraints. By incorporating prescribed-performance barriers into the min–max learning framework, ZG-RL-PP achieves robust tracking, disturbance attenuation, and prescribed transient and steady-state performance within a user-specified time. The recursive-margin construction further enables the performance constraints to be imposed on high-relative-degree tracking channels. Theoretical analysis proves the practical predefined-time convergence of the critic learning error and the uniform ultimate boundedness of the closed-loop system. Comparative simulations on two HFA model sets confirm that ZG-RL-PP provides improved tracking accuracy, constraint satisfaction, and learning performance over the no-barrier controller and the nominal IRL benchmark. Model uncertainty simulations confirm ZG-RL-PP maintains its performance facing different model uncertainties. Future work will further investigate actuator degradation to extend the applicability of the proposed learning-based control framework.

Author Contributions: Conceptualization, H.Z. and M.W.; methodology, H.Z. and J.M.; software, H.Z.; validation, H.Z. and C.P.; formal analysis, H.Z.; data curation, H.Z. and Y.Z.; writing—original draft preparation, H.Z.; writing—review and editing, C.P.; visualization, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Patil, M.; Hodges, D.; Cesnik, C. Nonlinear aeroelasticity and flight dynamics of high-altitude long-endurance aircraft. *J. Aircr.* **2001**, *38*, 88–94.
2. Peng, C.; Ma, J. Online integral reinforcement learning control for an uncertain highly flexible aircraft using state and output feedback. *Aerosp. Sci. Technol.* **2021**, *109*, 106442.
3. Gregory, I. Modified dynamic inversion to control large flexible aircraft: What’s going on? In Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit, Reston, VA, USA, 1999. AIAA Paper 1999-3998.
4. Patil, M.; Hodges, D. Output feedback control of the nonlinear aeroelastic response of a slender wing. *J. Guid. Control Dyn.* **2002**, *25*, 302–308.
5. Shearer, C.; Cesnik, C. Trajectory control for very flexible aircraft. *J. Guid. Control Dyn.* **2008**, *31*, 340–357.
6. Raghavan, B.; Patil, M. Flight control for flexible high-aspect-ratio flying wings. *J. Guid. Control Dyn.* **2010**, *33*, 64–74.
7. Dillsaver, M.; Cesnik, C.; Kolmanovsky, I. Trajectory control of very flexible aircraft with gust disturbance. In Proceedings of the AIAA Atmospheric Flight Mechanics Conference, Reston, VA, USA, 2013. AIAA Paper 2013-4745.
8. Cook, R.; Palacios, R.; Goulart, P. Robust gust alleviation and stabilization of very flexible aircraft. *AIAA J.* **2013**, *51*, 330–340.
9. Haghighat, S.; Liu, H.; Martins, J. Model-predictive gust load alleviation controller for a highly flexible aircraft. *J. Guid. Control Dyn.* **2012**, *35*, 1751–1766.
10. Bianchi, D.; Di Gennaro, S.; Di Ferdinando, M.; Acosta Lúa, C. Robust control of UAV with disturbances and uncertainty estimation. *Machines* **2023**, *11*, 352.
11. Liu, X.; Sun, Q.; Cooper, J. LQG based model predictive control for gust load alleviation. *Aerosp. Sci. Technol.* **2017**, *71*, 499–509.
12. Qu, Z.; Annaswamy, A. Adaptive output-feedback control with closed-loop reference models for very flexible aircraft. *J. Guid. Control Dyn.* **2016**, *39*, 873–888.
13. Lewis, F.; Vrabie, D. Reinforcement learning and adaptive dynamic programming for feedback control. *IEEE Circuits Syst. Mag.* **2009**, *9*, 32–50.
14. Lewis, F.; Vrabie, D.; Vamvoudakis, K. Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers. *IEEE Control Syst. Mag.* **2012**, *32*, 76–105.

15. Kiumarsi, B.; Vamvoudakis, K.; Modares, H.; Lewis, F. Optimal and autonomous control using reinforcement learning: A survey. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 2042–2062.
16. Vrabie, D.; Pastravanu, O.; Abu-Khalaf, M.; Lewis, F. Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica* **2009**, *45*, 477–484.
17. Jiang, Y.; Jiang, Z.P. Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. *Automatica* **2012**, *48*, 2699–2704.
18. Zhu, L.; Modares, H.; Peen, G.; Lewis, F.; Yue, B. Adaptive suboptimal output-feedback control for linear systems using integral reinforcement learning. *IEEE Trans. Control Syst. Technol.* **2015**, *23*, 264–273.
19. Abu-Khalaf, M.; Lewis, F.; Huang, J. Neurodynamic programming and zero-sum games for constrained control systems. *IEEE Trans. Neural Netw.* **2008**, *19*, 1243–1252.
20. Vamvoudakis, K.; Lewis, F. Online solution of nonlinear two-player zero-sum games using synchronous policy iteration. *Int. J. Robust Nonlinear Control* **2012**, *22*, 1460–1483.
21. Wang, P.; Yu, C.; Lv, M.; Duan, G.R. Safe fixed-time reinforcement learning for nonlinear zero-sum games with obstacle avoidance awareness. *Automatica* **2026**, *183*, 112673.
22. Yuan, X.; Chen, B.; Lin, C. Neural adaptive fixed-time control for nonlinear systems with full-state constraints. *IEEE Trans. Cybern.* **2023**, *53*, 3048–3059.
23. Bechlioulis, C.; Rovithakis, G. Prescribed performance adaptive control for SISO feedback linearizable nonlinear systems. *Automatica* **2008**, *44*, 1858–1865.
24. Bechlioulis, C.; Rovithakis, G. Robust adaptive control of feedback linearizable MIMO nonlinear systems with prescribed performance. *IEEE Trans. Autom. Control* **2009**, *54*, 2090–2099.
25. Bechlioulis, C.; Rovithakis, G. A low-complexity global approximation-free control scheme with prescribed performance for unknown pure feedback systems. *Automatica* **2014**, *50*, 1217–1226.
26. Song, Y.; Wang, Y.; Holloway, J.; Krstic, M. Time-varying feedback for regulation of normal-form nonlinear systems in prescribed finite time. *Automatica* **2017**, *83*, 243–251.
27. Holloway, J.; Krstic, M. Prescribed-time observers for linear systems in observer canonical form. *IEEE Trans. Autom. Control* **2019**, *64*, 3905–3912.
28. Sánchez-Torres, J.; Gómez-Gutiérrez, D.; López, E.; Loukianov, A. A class of predefined-time stable dynamical systems. *IMA J. Math. Control Inf.* **2018**, *35*, i1–i29.
29. Wang, H.; Shi, S.; Zhen, Z. Flexible performance-based predefined-time tracking control for unmanned helicopter with irregular constraints. *Aerosp. Sci. Technol.* **2025**, *164*, 110445.
30. Mao, B.; Wu, X.; Liu, H.; Xu, Y.; Lü, J. Adaptive fuzzy tracking control with global prescribed-time prescribed performance for uncertain strict-feedback nonlinear systems. *arXiv* **2022**, *arXiv:2210.16477*.
31. Wang, Q.; Cao, J.; Liu, H. Adaptive fuzzy control of nonlinear systems with predefined time and accuracy. *IEEE Trans. Fuzzy Syst.* **2022**, *30*, 5152–5165.
32. Ni, J.; Shi, P. Global predefined time and accuracy adaptive neural network control for uncertain strict-feedback systems with output constraint and dead zone. *IEEE Trans. Syst. Man Cybern. Syst.* **2021**, *51*, 7903–7918.
33. Gibson, T.; Annaswamy, A.; Lavretsky, E. Modeling for Control of Very Flexible Aircraft. In Proceedings of the AIAA Guidance, Navigation, and Control Conference, 2011. AIAA 2011-6202.
34. Ma, J.; Peng, C. Adaptive model-free fault-tolerant control based on integral reinforcement learning for a highly flexible aircraft with actuator faults. *Aerosp. Sci. Technol.* **2021**, *119*, 107204.
35. Qu, Z.; Thomsen, B.; Annaswamy, A. Adaptive control for a class of multi-input multi-output plants with arbitrary relative degree. *IEEE Trans. Autom. Control* **2020**, *65*, 3023–3038.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.